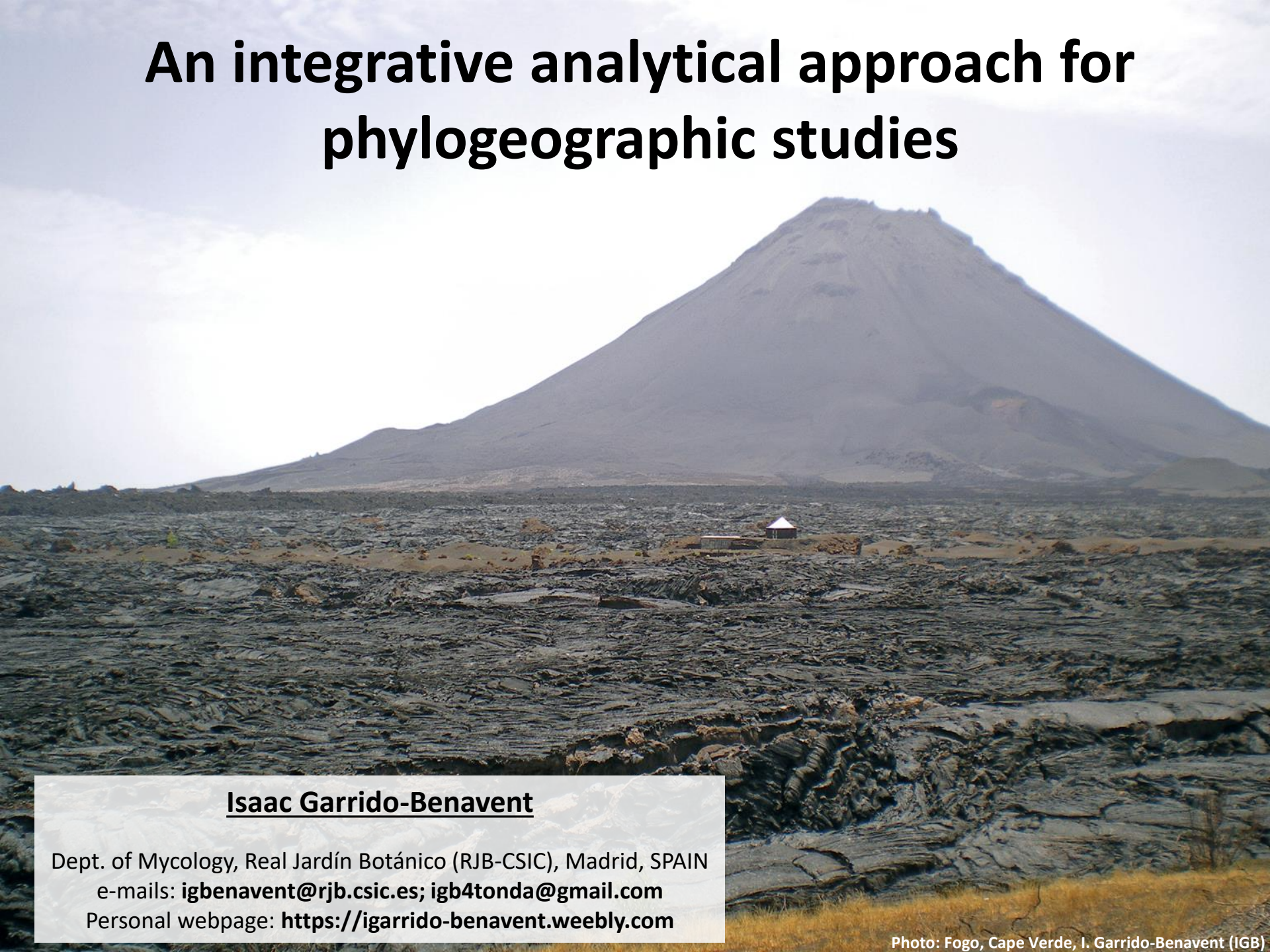
An integrative analytical approach for phylogeographic studies



Isaac Garrido-Benavent

Dept. of Mycology, Real Jardín Botánico (RJB-CSIC), Madrid, SPAIN
e-mails: **igbenavent@rjb.csic.es; igb4tonda@gmail.com**
Personal webpage: **https://igarrido-benavent.weebly.com**

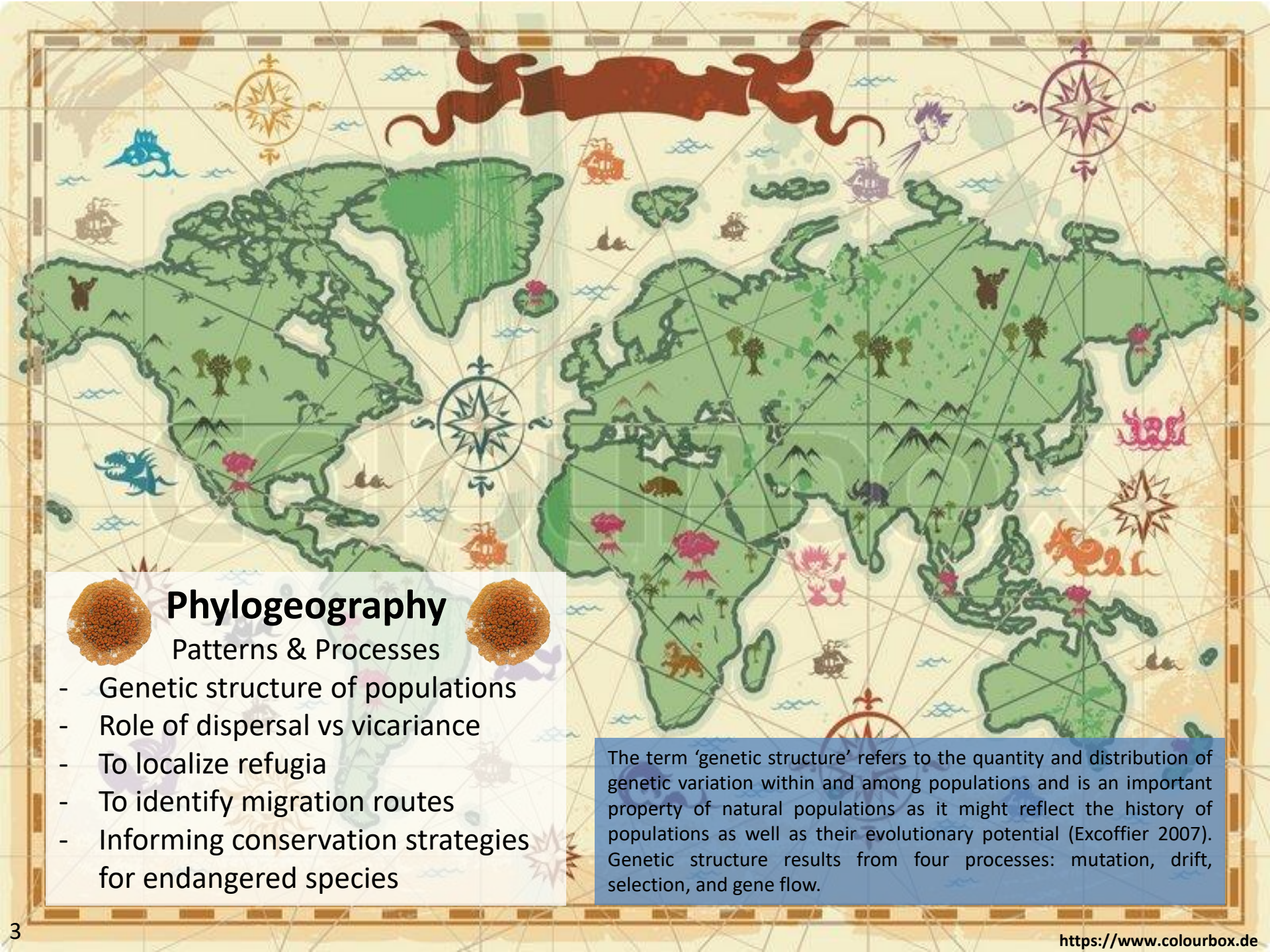
Photo: Fogo, Cape Verde, I. Garrido-Benavent (IGB)

"…a field of study concerned with the principles and processes governing the geographical distributions of genealogical lineages, especially those within and among closely related species…" (Avise 2000)

Phylogenetics **Population genetics**

Phylogeography (Avise et al. 1987)

Photo: Sal, Cape Verde, IGB



Phylogeography

Patterns & Processes

- Genetic structure of populations
- Role of dispersal vs vicariance
- To localize refugia
- To identify migration routes
- Informing conservation strategies for endangered species

The term ‘genetic structure’ refers to the quantity and distribution of genetic variation within and among populations and is an important property of natural populations as it might reflect the history of populations as well as their evolutionary potential (Excoffier 2007). Genetic structure results from four processes: mutation, drift, selection, and gene flow.

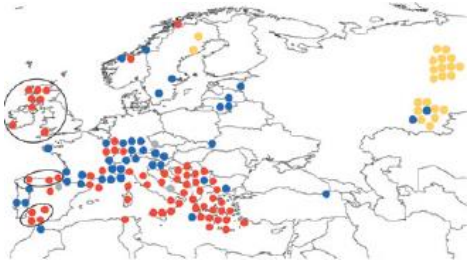
Guidelines for conducting a phylogeographic study

Brief introduction to RevBayes software

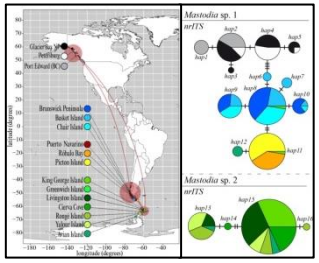
Brief tutorials for particular analyses

Delineating the scope of a phylogeographic study

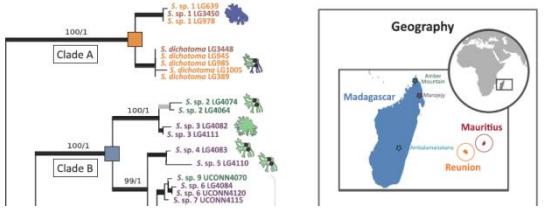
Taxonomic complexity



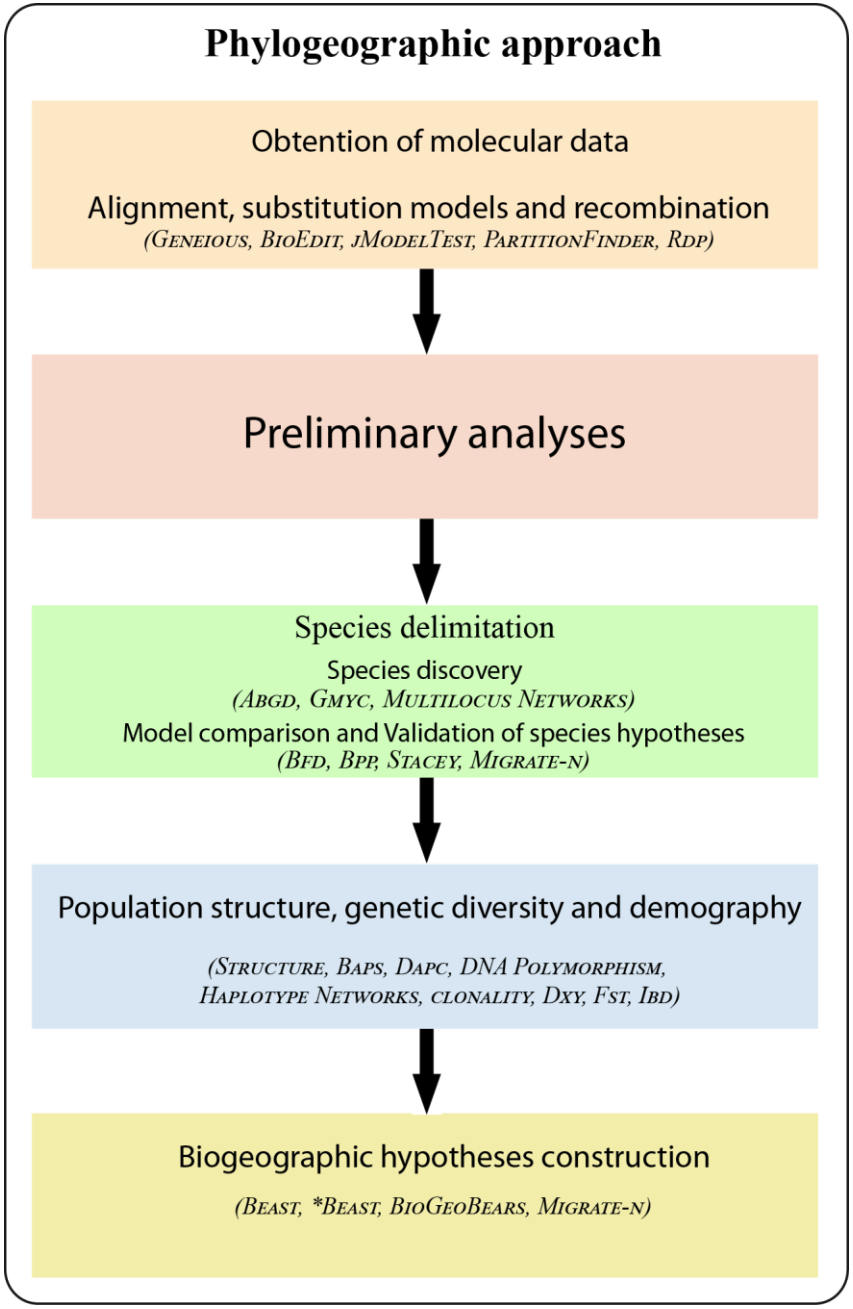
1 species
(e.g. *Lobaria pulmonaria*, Widmer et al. 2012)

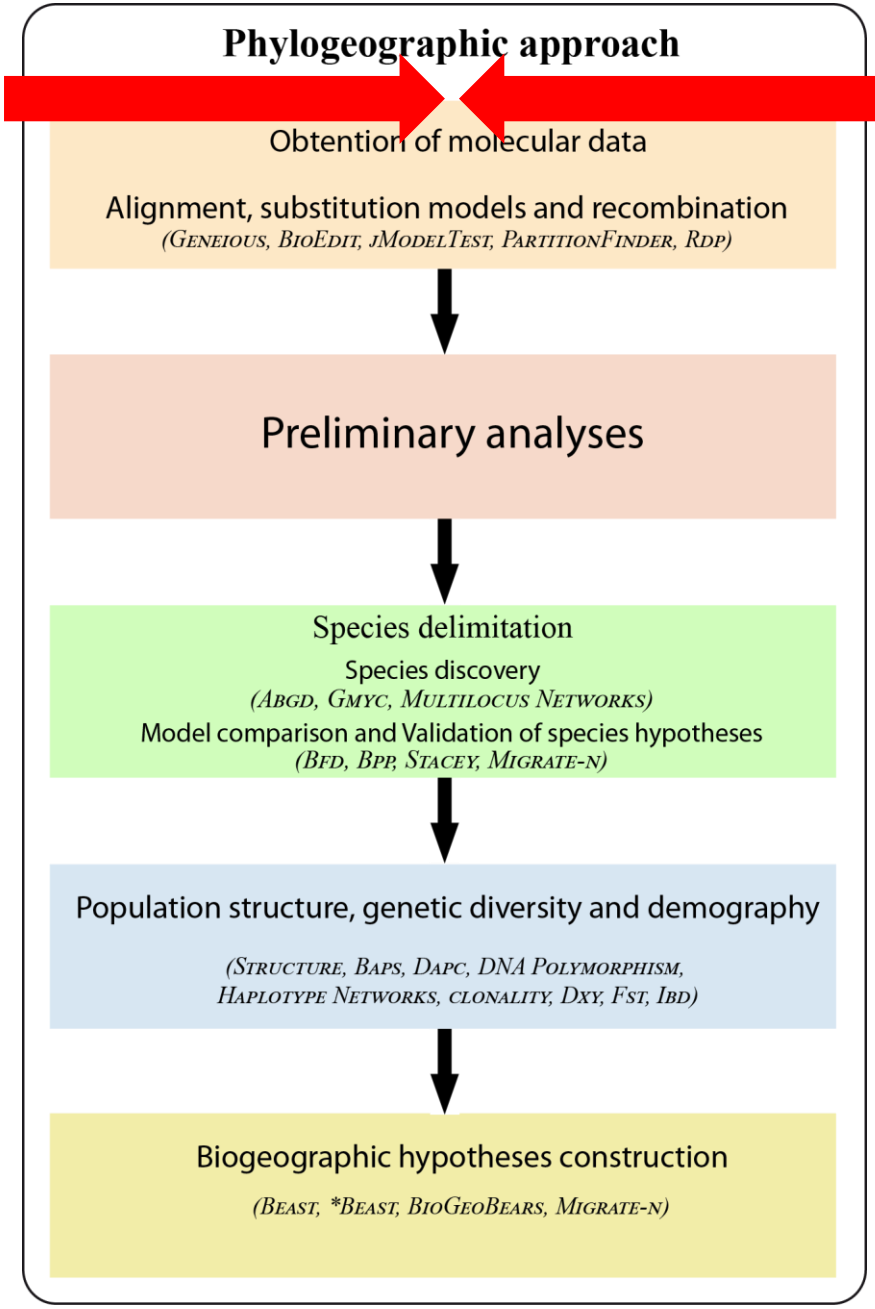


Two or a few closely related species
(e.g. *Mastodia tessellata s.l.*, Garrido-Benavent et al. 2018)



Several to many species within a genus
(e.g. *Sticta*, Simon et al. 2018)


Phylogeographic approach
Obtention of molecular data
Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)
Preliminary analyses
Species delimitation
Species discovery
(Abgd, Gmyc, Multilocus Networks)
Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)
Population structure, genetic diversity and demography
(Structure, Baps, Dapc, DNA Polymorphism,
Haplotype Networks, clonality, Dxy, Fst, Ibd)
Biogeographic hypotheses construction
(Beast, *Beast, BioGeoBears, Migrate-n)



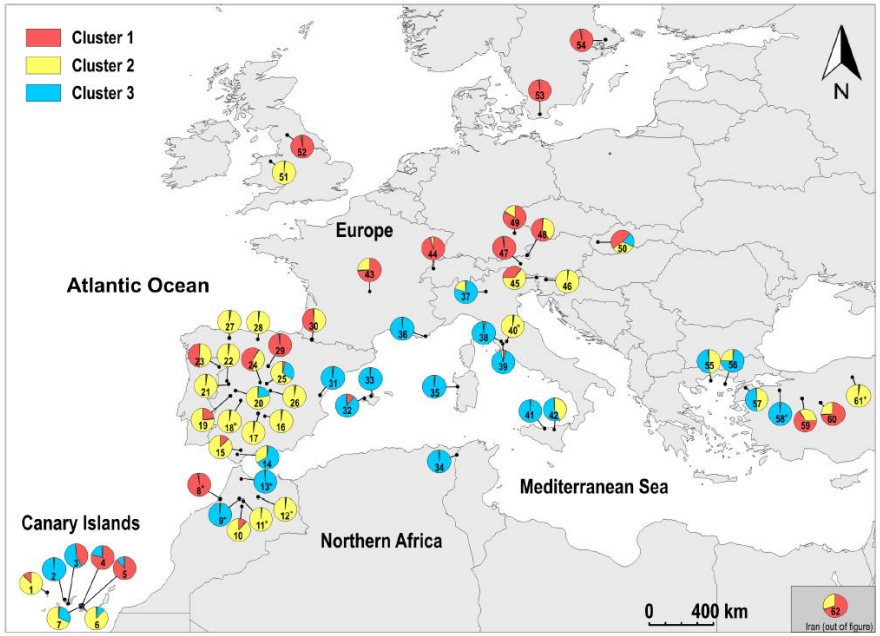
Sampling design

The first topic we must care about is sampling. Its design should be consciously prepared according to the hypotheses we want to test and taking advantage of our experience in previous phylogeographic works.

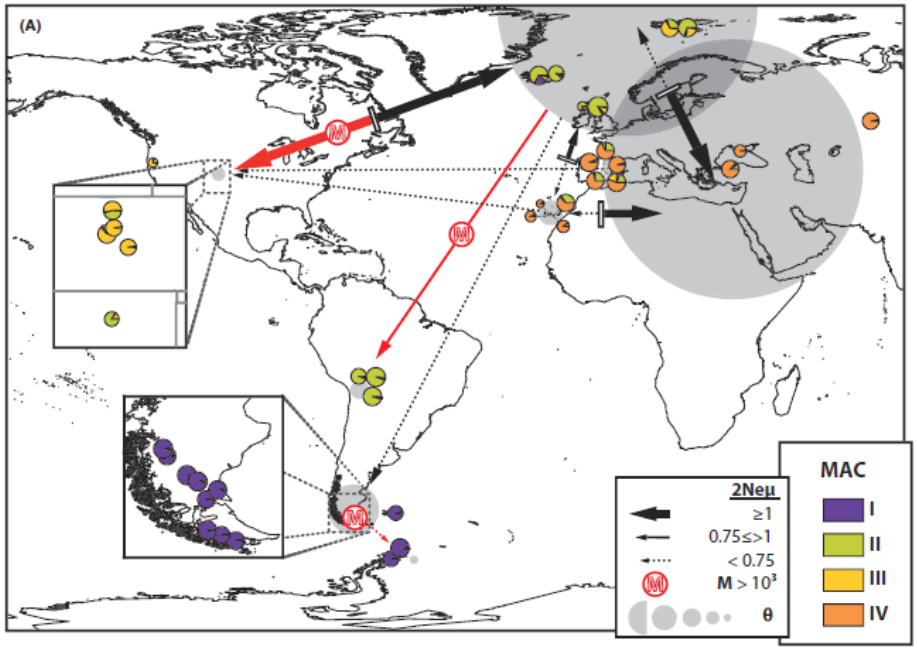
The robustness of biogeographic inferences depends on the original sampling design

The higher number of **localities**, the better.

The higher number of **specimens**, the better.



62 localities of *Parmelina tiliacea* representing different altitudes and habitats (Adapted from Núñez-Zapata et al. 2015)



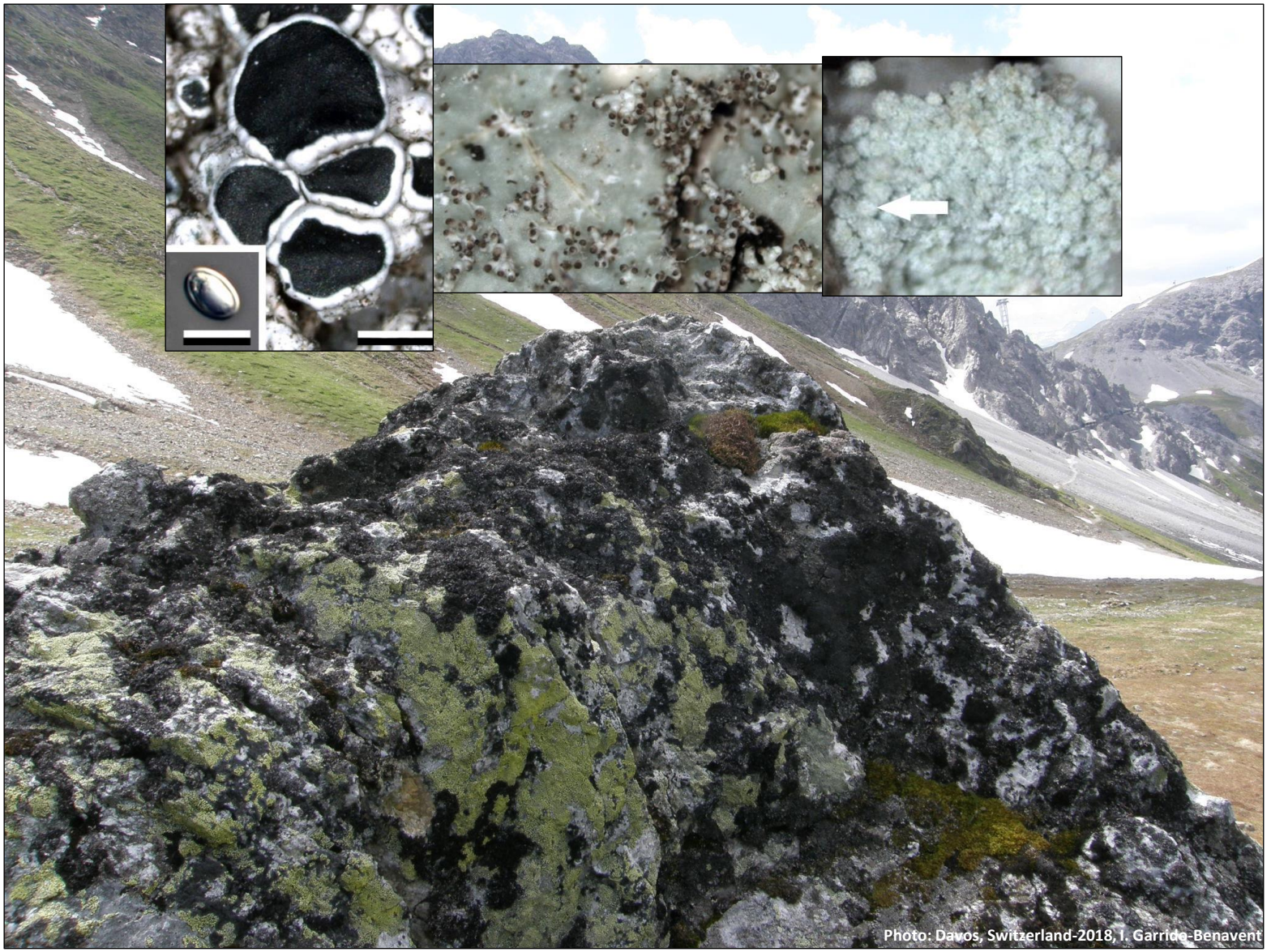
39 localities from seven regions in *Cetraria aculeata* (Adapted from Fernández-Mendoza & Printzen 2013)

364 specimens (1-14 spec. /locality)

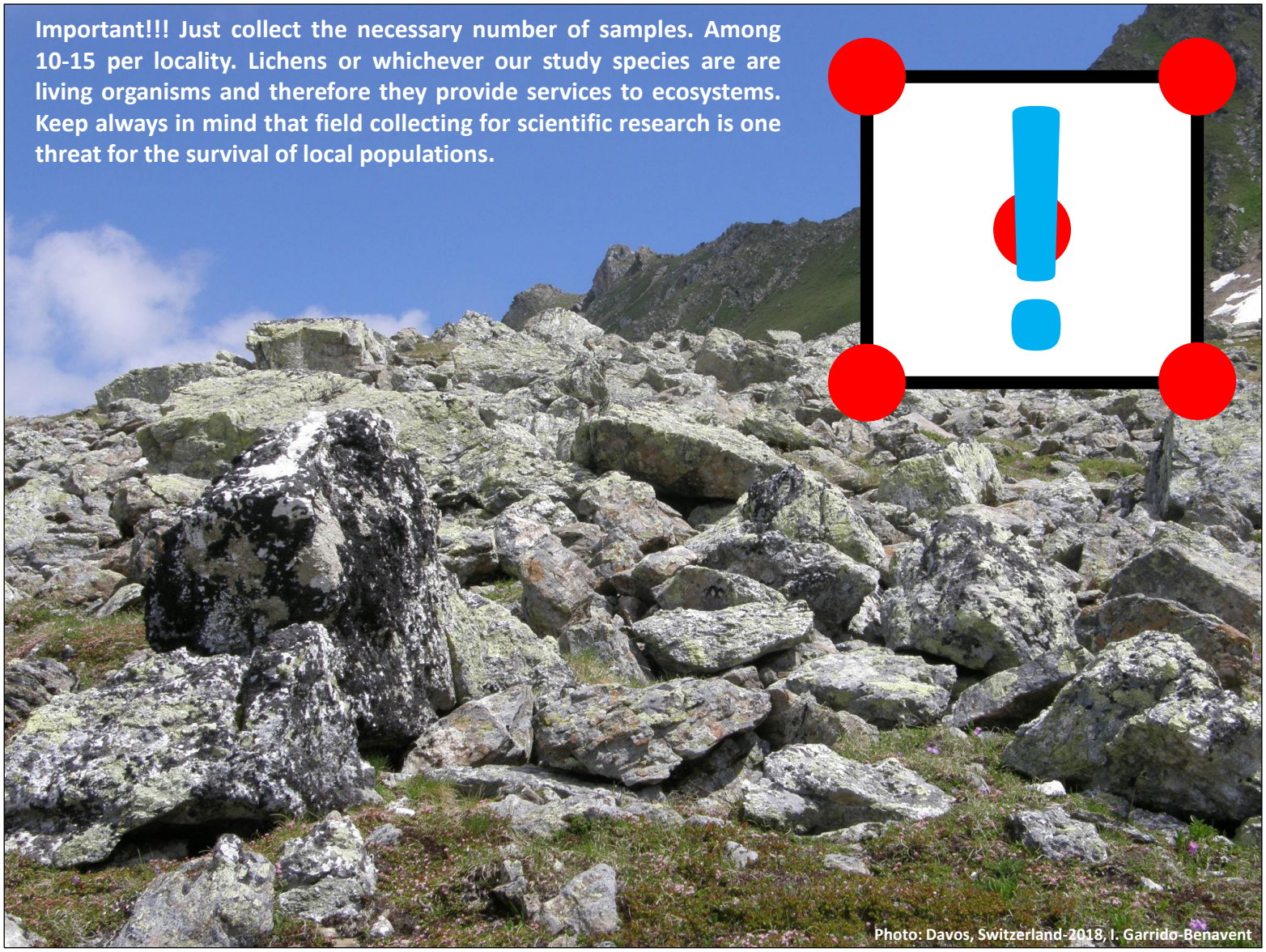
356 specimens (to 12 spec./locality)

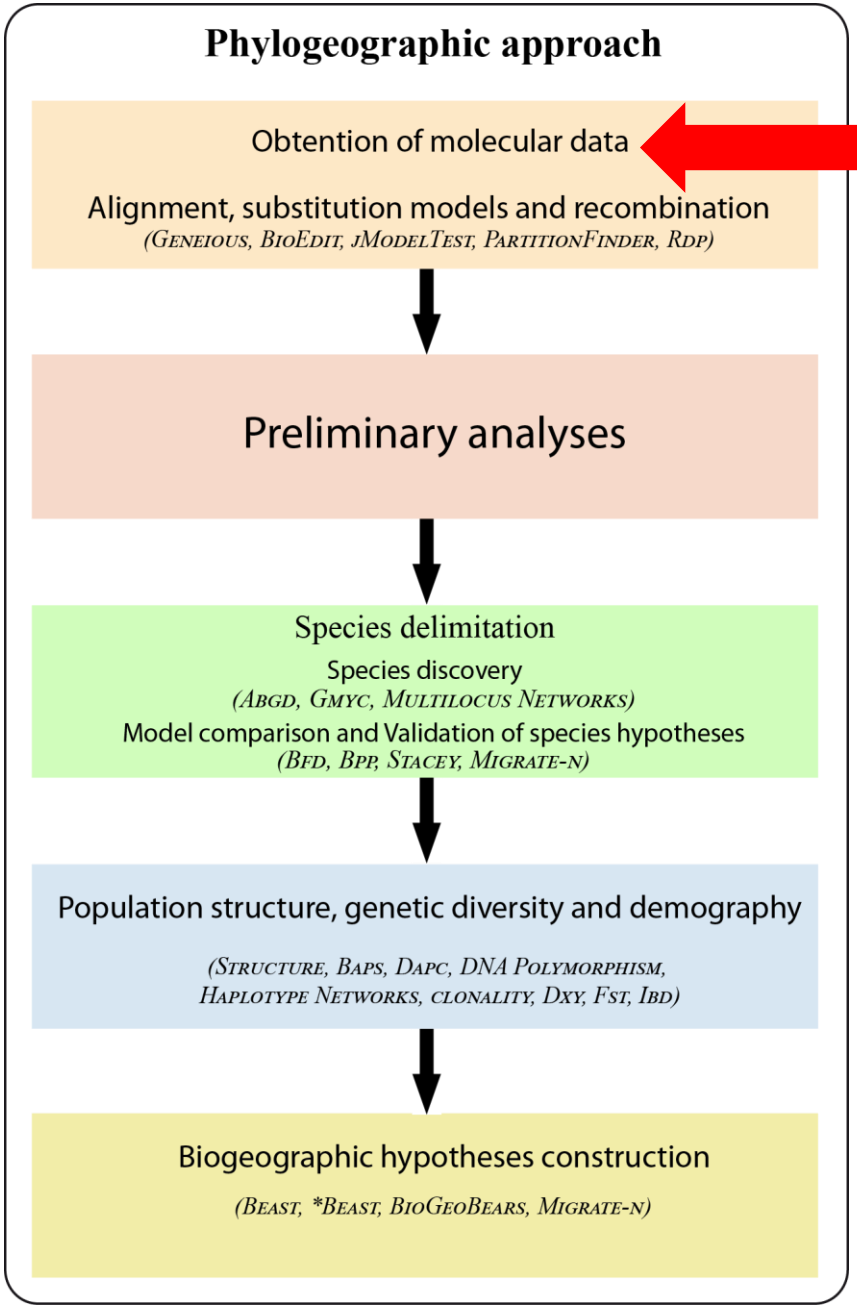
4323 specimens in **142** localities from Europe in *Lobaria pulmonaria* (Widmer et al. 2012)

The reproductive mode of the study species must be taken into account when sampling



Take samples at different places within a sampling locality





Molecular data

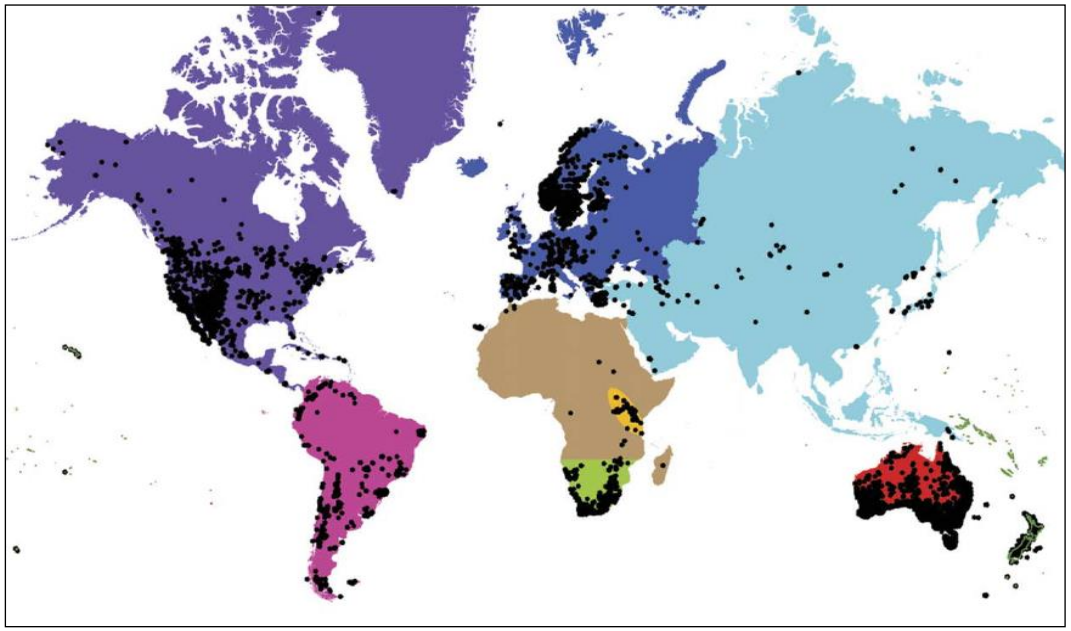
A) DNA sequences from one to X genomic regions

B) Simple sequence repeat (SSR) markers (e.g. microsatellites)

C) Restriction associated DNA sequencing (RADseq) or other type of data obtained through high-throughput sequencing methods

DNA sequences are appropriate for assessing historical biogeography in lichens

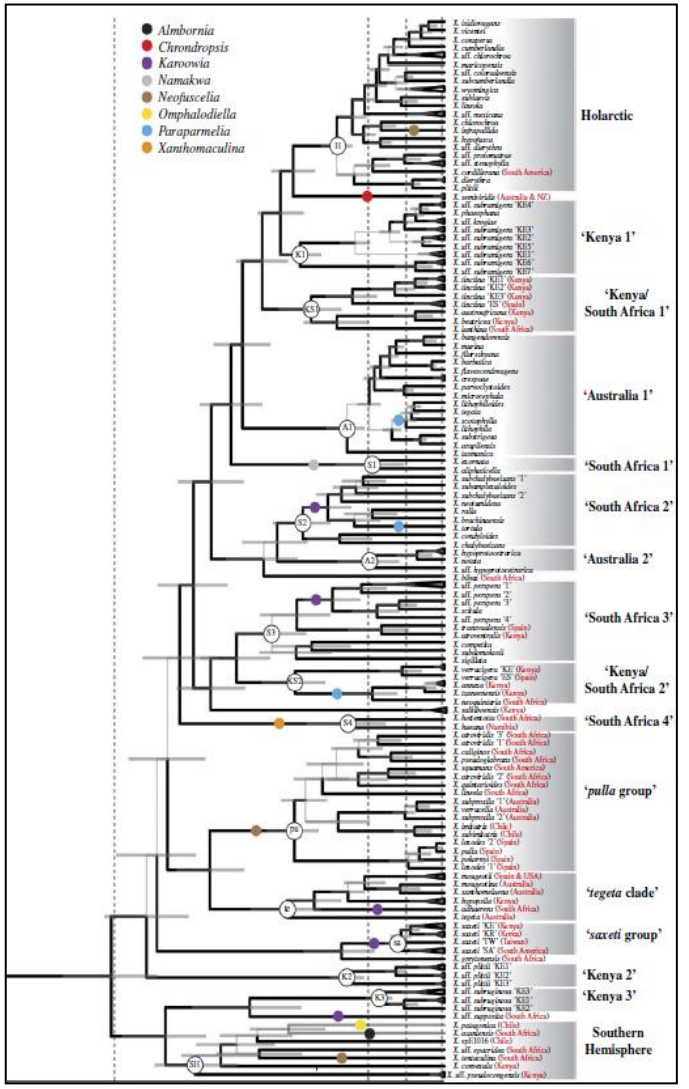
Genus *Xanthoparmelia* (*Parmeliaceae*)



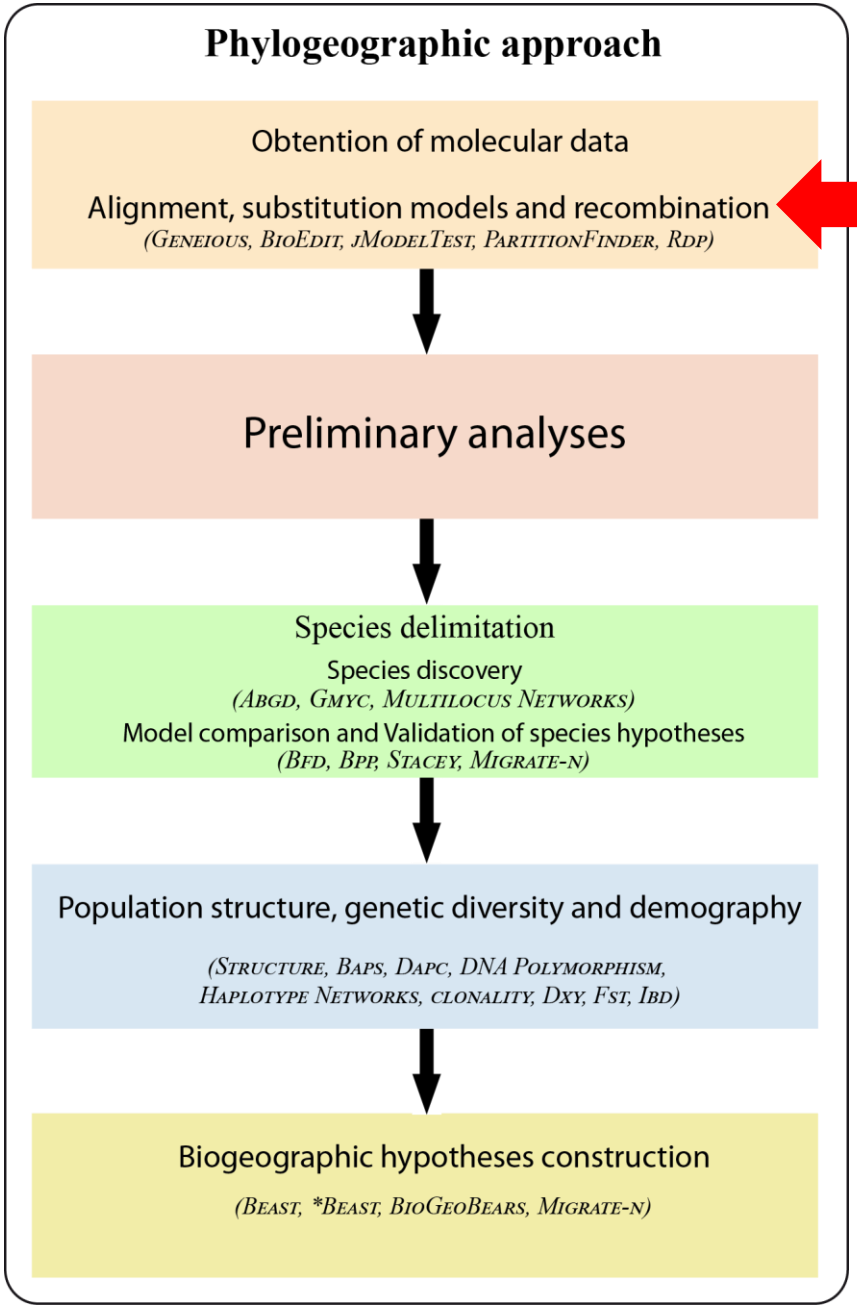
Sanger sequencing

Metagenomic reads

9 markers



Adapted from Leavitt et al. (2018)



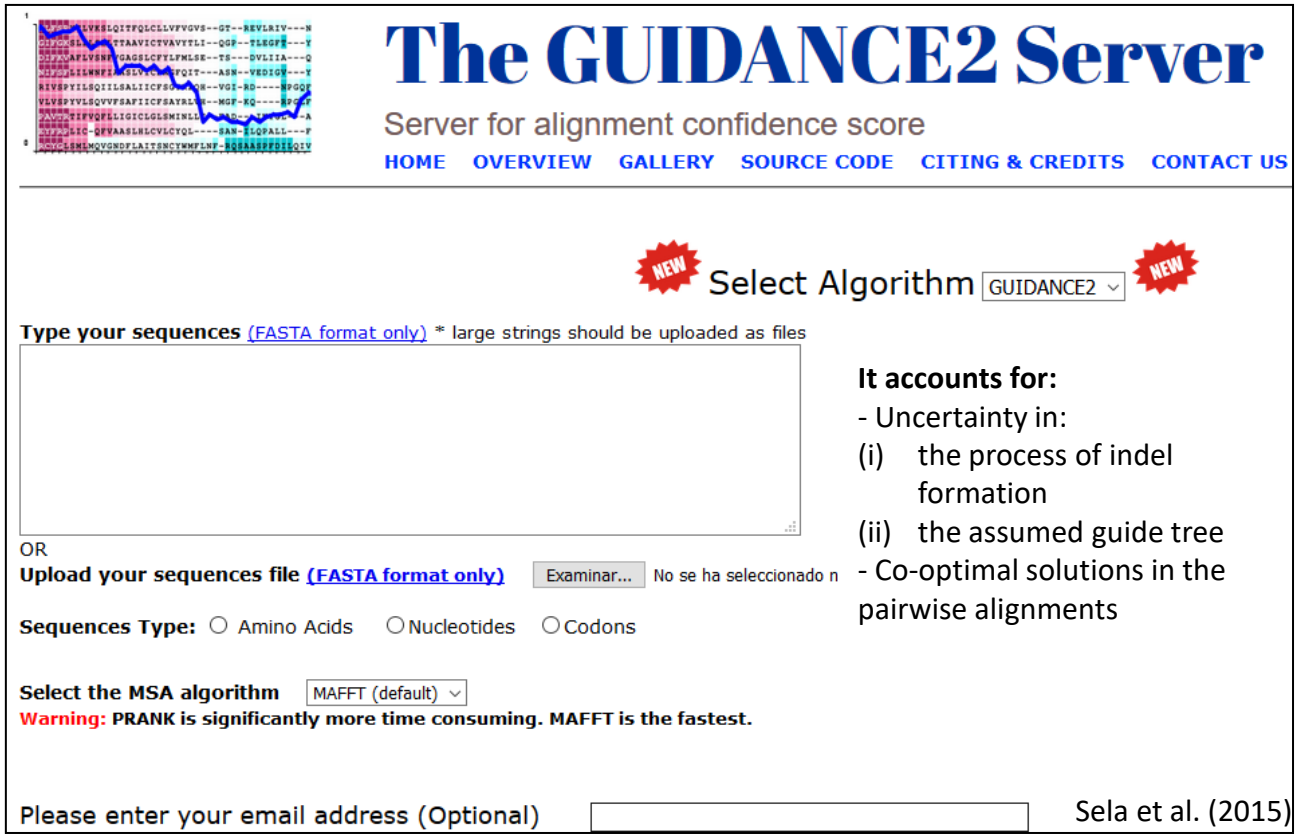
Alignment with MAFFT

Katoh & Standley (2013)

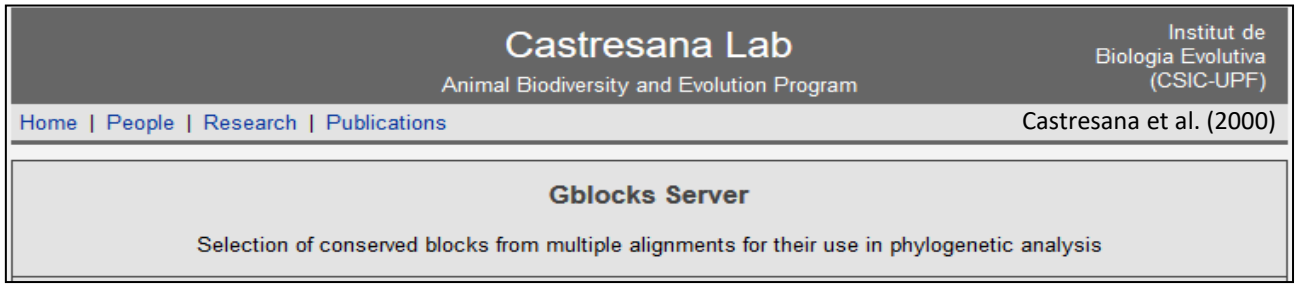
Once we assemble a DNA sequence dataset, the next step is to produce an alignment. This is one of the most crucial steps in the phylogeographic approach because all our next analyses will depend upon the accuracy of the alignment. There is a number of available tools to align DNA sequences, and a very rich literature discussing the pros and cons of each method. It has been found that MAFFT is both computationally efficient and has relatively high performance on both simulated and empirical data sets. The different options available in MAFFT work well for fungal and algal DNA sequences. It can be installed on your personal computers, run online or is available within other software such as Geneious.

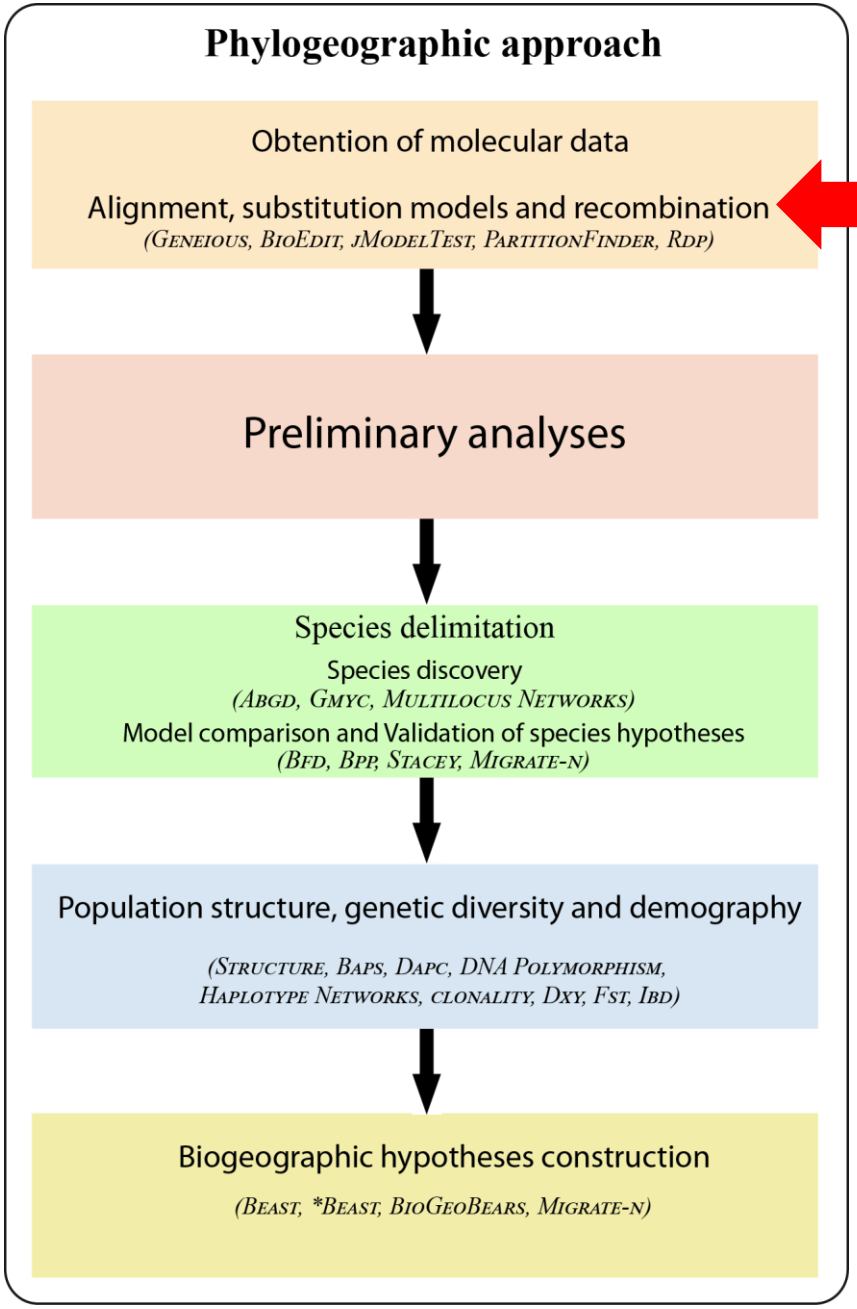
https://mafft.cbrc.jp/alignment/software/

GUIDANCE2 – to identify unreliable alignment regions



GBlocks – to automatically deal with gappy regions in complicated alignments (e.g. ITS)





For a particular dataset, inference of substitution models should be carried out using **the same** framework (ML or Bayesian) under which phylogenetic trees will be estimated

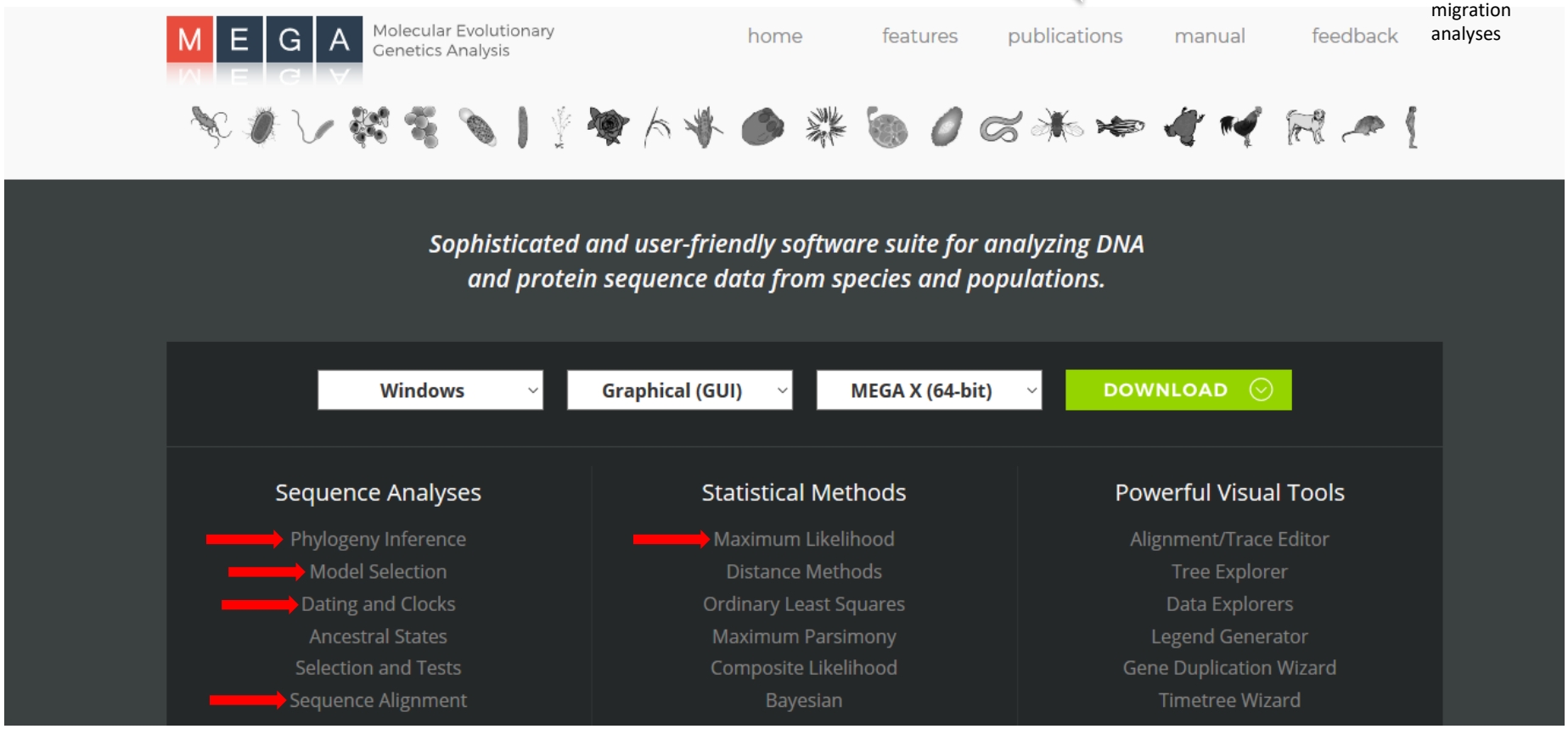
Next, it is important to estimate a substitution model for each DNA alignment. This information will feed further analyses in phylogeography.

MEGA Molecular Evolutionary Genetics Analysis

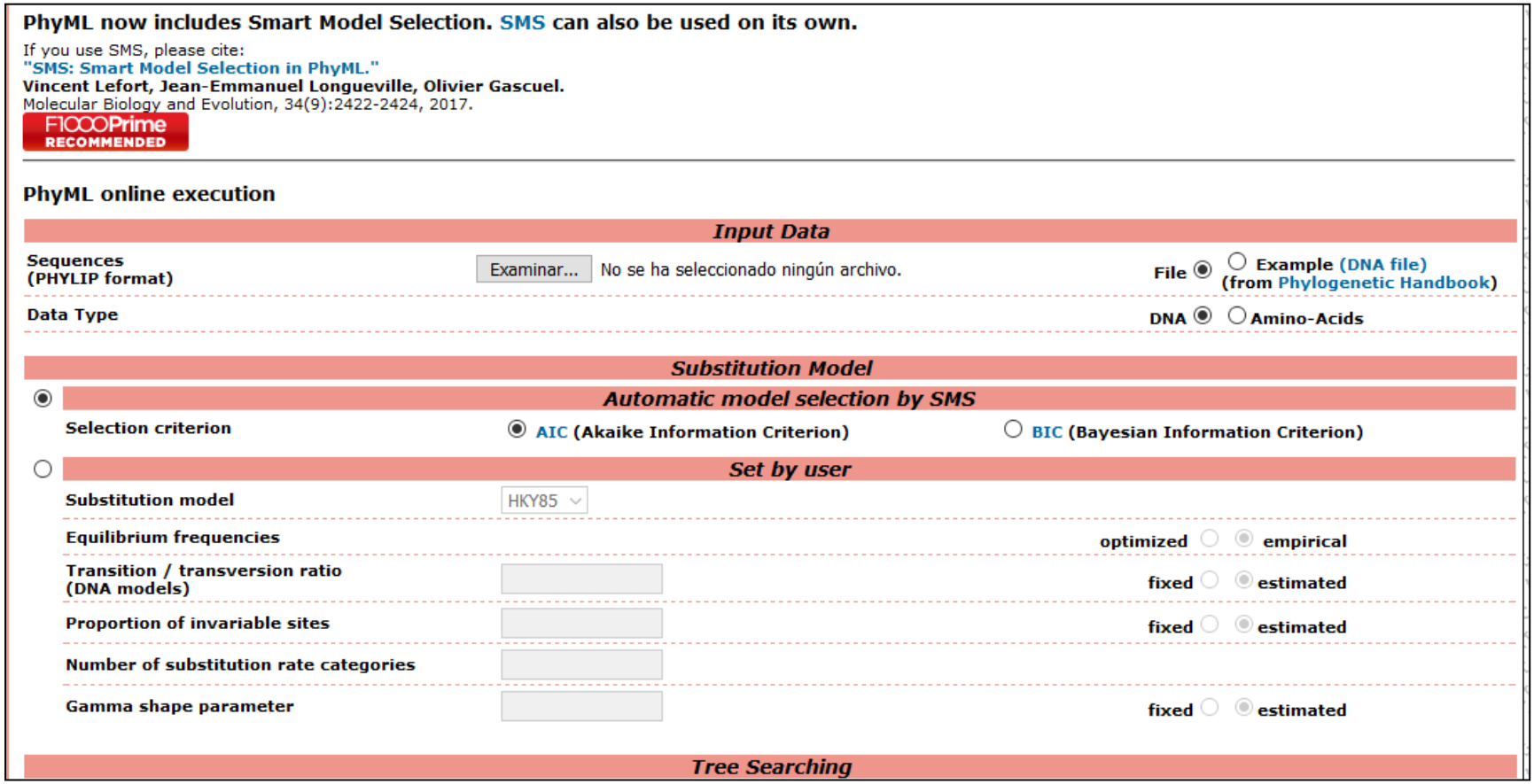
Preliminary phylogenies in MEGA (Kumar et al. 2018)

- DNA sequences alignment
- Estimation of the best substitution model
- Inference of phylogenies under a ML framework
- Calculating Transition/Transversion rates
- Molecular clock tests

Important for species delimitation, dating and migration analyses



Estimation of best substitution models and phylogenies under a ML framework



JMODELTEST

Darriba et al. (2012)

RAxML

Our standard tool for Maximum-likelihood based phylogenetic inference.

Stamatakis (2006); Stamatakis et al. (2008)

Estimation of best substitution models and phylogenies under a Bayesian framework

JMODELTEST

RAxML

Our standard tool for Maximum-likelihood based phylogenetic inference.

Mr Bayes

Bayesian Inference of Phylogeny

Ronquist et al. (2012)

PartitionFinder

Lanfear et al. (2012)

The best option is to:

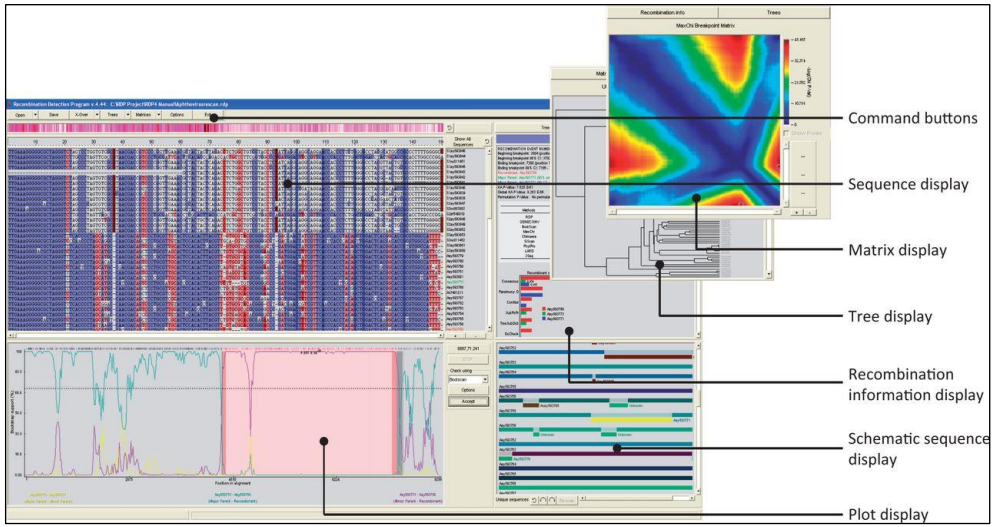
Estimate alternative phylogenies using different partitioning schemes and substitution models and compare marginal likelihoods (Model Comparison)

BEAST

Bayesian Evolutionary Analysis Sampling Trees

Drummond et al. (2012)

Detection of recombination

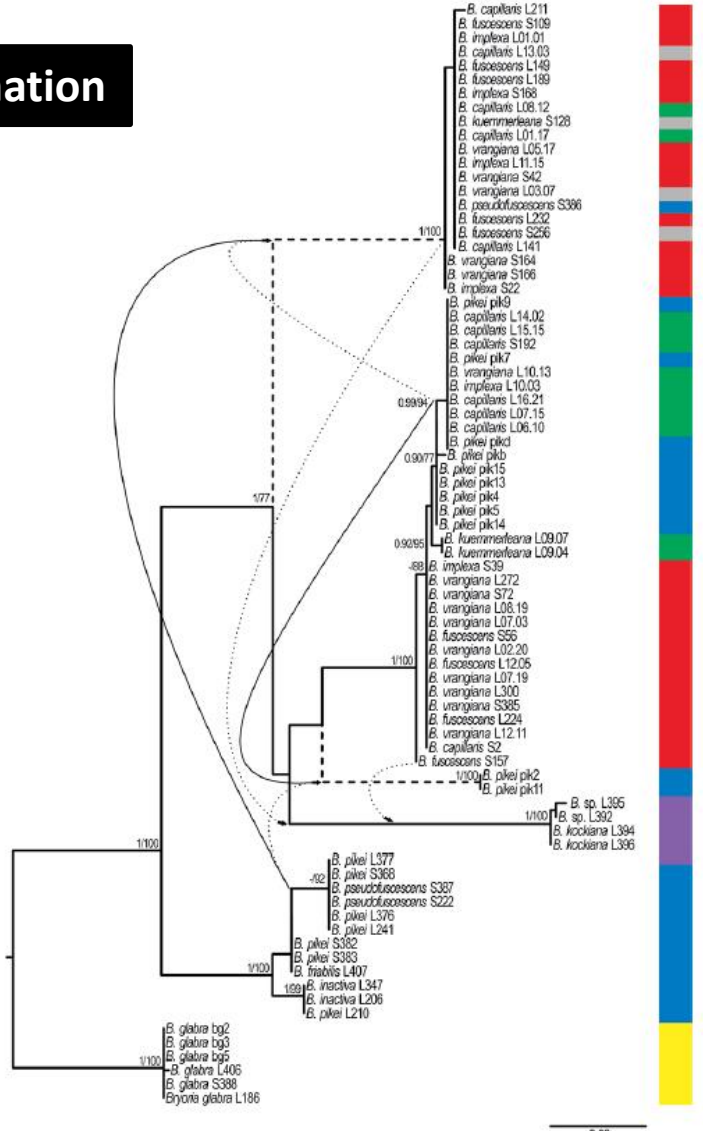


Adapted from Martin et al. (2015)

RDP4 software (Martin et al. 2010)

- RDP (Martin & Rybicki 2000)
- GENECONV (Padidam et al. 1999)
- MaxChi (Maynard-Smith 1992)

Fungal ITS→high polymorphism→Spurious recombination signal

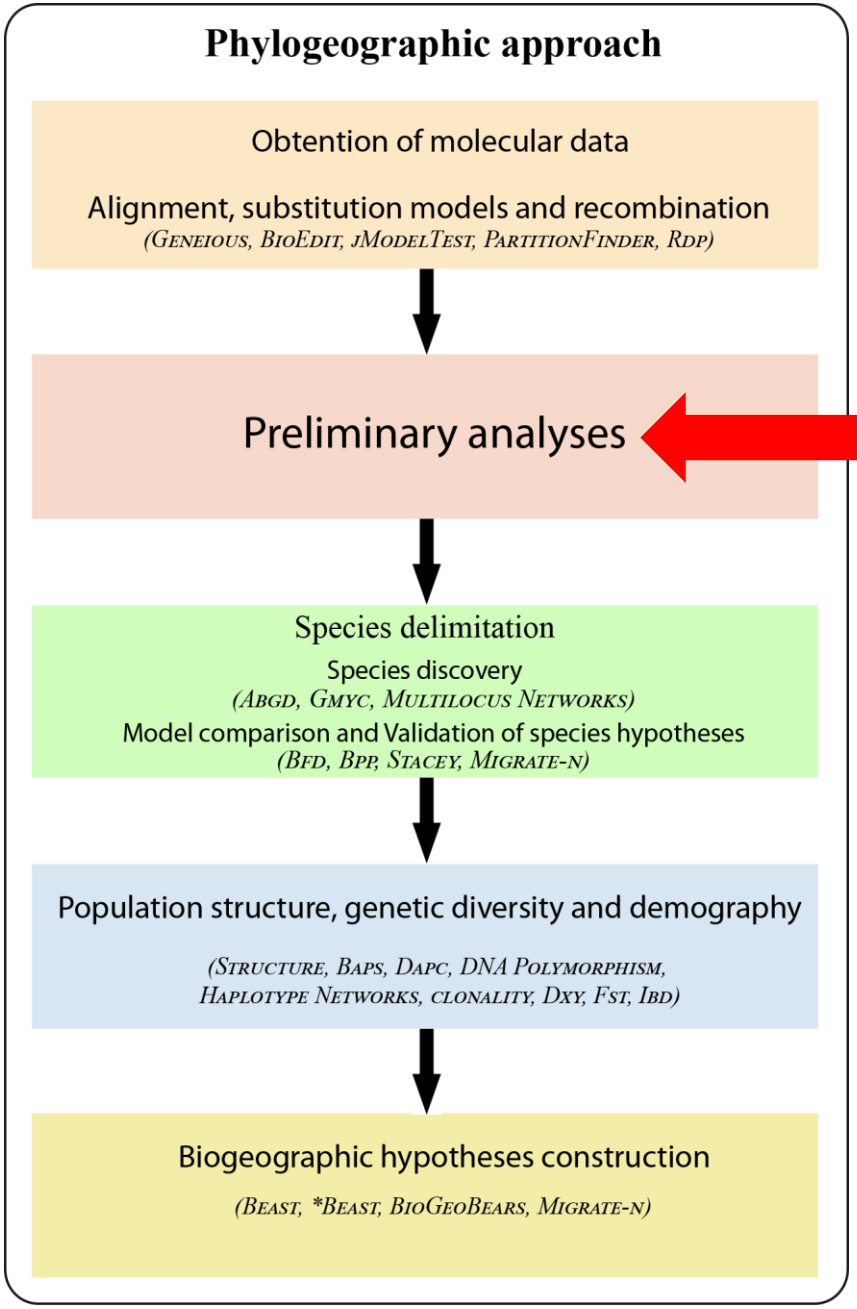


Adapted from Boluda et al. (2018)

Assessing topological conflicts among loci

Congruence among distance matrices (CADM) test (Legendre & Lapointe 2004, Campbell et al. 2011)

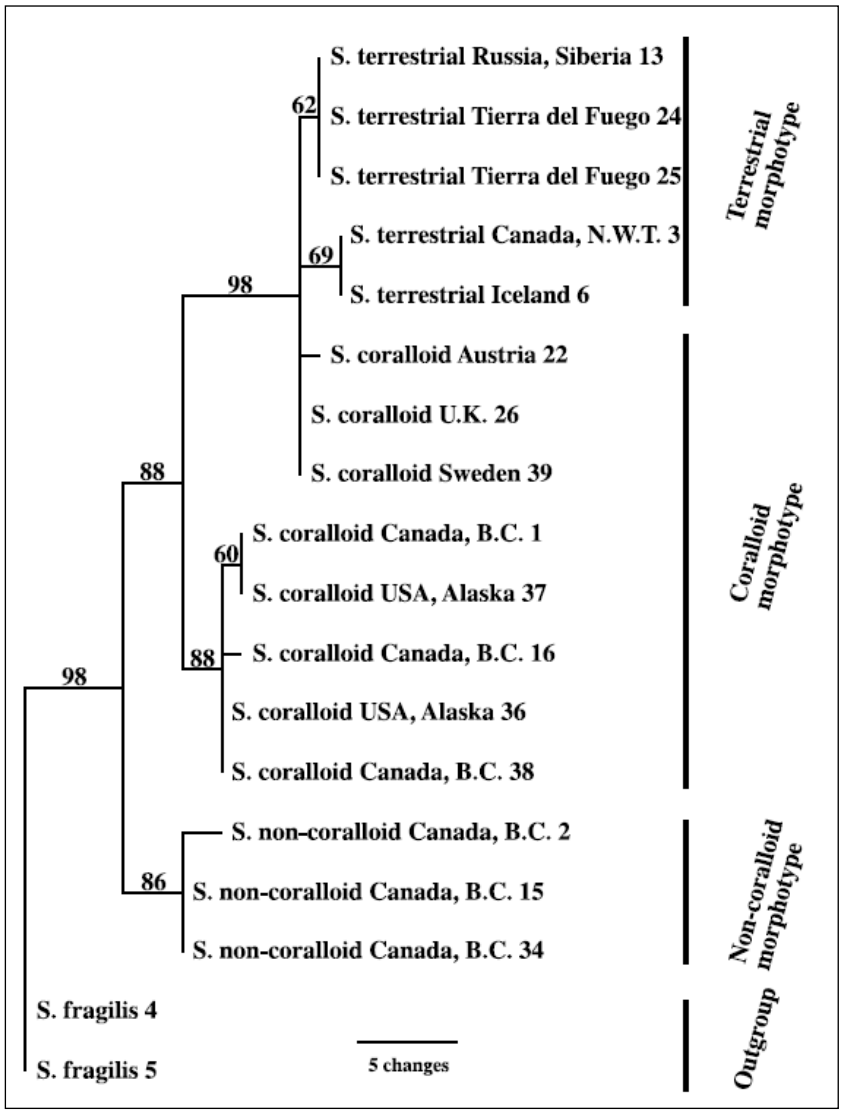
CADM.global function implemented in the library "*ape*" (Paradis et al. 2004) in R



To assess species boundaries in our study spp.

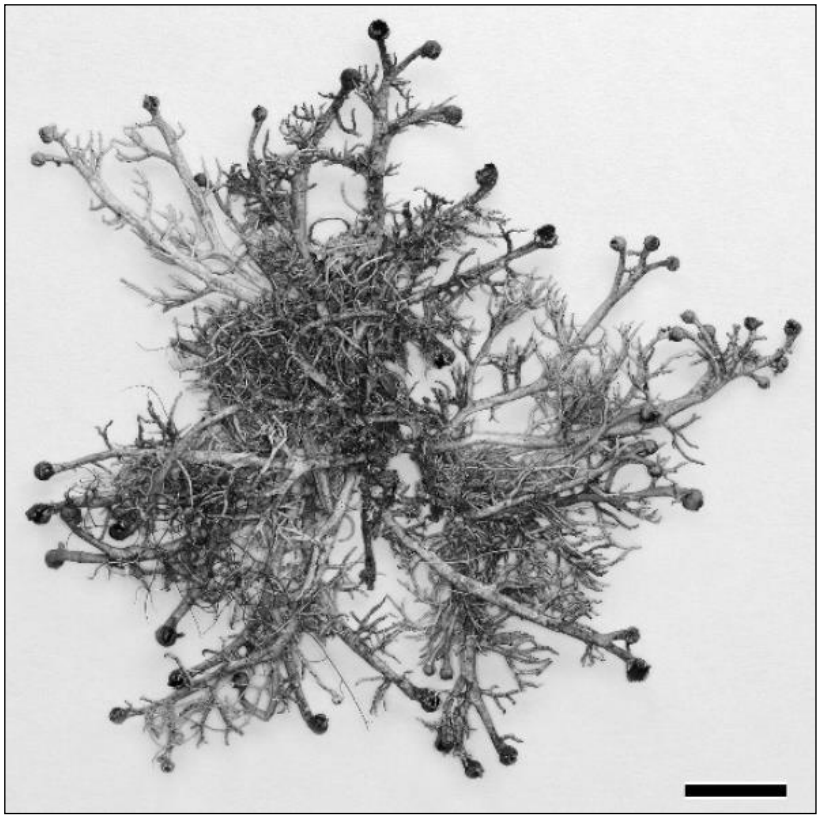
The next step in any phylogeographic study in which we include a huge number of specimens from widely distant localities is to detect whether there are evidences of cryptic speciation. This is, whether we have more than one species in our dataset. It is especially important to treat each putative species separately because equivocal concepts of species limits could confound an extrinsic barrier to dispersal with intrinsic reproductive barriers (Pante et al., 2015).

Phylogeographic studies may contribute to unravel new species



Phylogeny of the *Sphaerophorus globosus* species complex using 5 genes and inferred under a parsimony framework in PAUP* (Högnabba & Wedin 2003)

In fact, phylogeographic studies have been pivotal for the discovery of many new taxa for science. In this study by Hognabba and Wedin (2003) on the *Sphaerophorus globosus* species complex, they revealed two phylogenetic species, one restricted to hyperoceanic areas along the North American Pacific Northwest, subsequently described as *S. venerabilis* (Wedin et al., 2009), and the second displaying a wide distribution in both hemispheres.



Holotype of *S. venerabilis*. Scale = 1 cm (Wedin et al. 2009)

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(Abgd, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

Biogeographic hypotheses construction

*(Beast, *Beast, BioGeoBears, Migrate-n)*

Integrative taxonomy

Literature search

Macroscopic and microscopic data

Phylogenetic relationships
(RAxML, MrBayes, Mega, PhyML)

Chemical compounds

A series of preliminary analyses are needed to ascertain how many species we are working with. Once we find out evidence of the existence of cryptic species in our dataset, for example constructing preliminary phylogenies, we can take a look at the literature or also conduct some analyses of traditional taxonomy to test whether there are further evidences of more than one species. Integrating traditional and molecular-based methods is an approach that has been named "Integrative taxonomy", which is widely used now in lichenology.

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(Abgd, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

Biogeographic hypotheses construction

*(Beast, *Beast, BioGeoBears, Migrate-n)*

Integrative taxonomy

Macroscopy and microscopic data

Literature search

Chemical compounds

DNA polymorphism
(DnaSP)

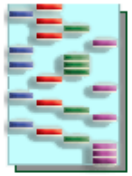
Phylogenetic relationships
(RAxML, MrBayes, Mega, PhyML)

Ecology
(Maxent, Wallace, hypervolume, ecospat, & ENMTools in R)

Age estimates
*(Beast, *Beast)*

Furthermore, this integrative taxonomy can also benefit from other sources of information. For example, we can calculate indices of DNA polymorphism and, after comparing with previous, similar works, determine whether they fit better scenarios of more than one species.

Calculation of DNA polymorphism indices with DnaSP



DnaSP DNA Sequence Polymorphism

Julio Rozas, Albert Ferrer-Mata, Juan Carlos Sánchez-DelBarrio, Sara Guirao-Rico, Pablo Librado, Sebastián Ramos-Onsins and Alejandro Sánchez-Gracia
Universitat de Barcelona
Current Beta Version: **6.12.01** (July 25, 2018)

DnaSP, DNA Sequence Polymorphism, is a software package for the analysis of DNA polymorphisms using data from a single locus (a multiple sequence aligned -MSA data), or from several loci (a Multiple-MSA data, such as formats generated by some assembler RAD-seq software). DnaSP can estimate several measures of DNA sequence variation within and between populations in noncoding, synonymous or nonsynonymous sites, or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow and gene conversion parameters. Moreover, DnaSP can conduct a number of neutrality tests, such as (among other), the Hudson, Kreitman and Aguadé (1987), Tajima (1989), McDonald and Kreitman (1991), Fu and Li (1993), and Fu (1997), Ramos-Onsins and Rozas (2002), Achaz (2009) tests, and compute their confidence intervals by the coalescent. The results of the analyses are displayed on tabular and graphic form.

Input Data Files:

- A single MSA (one genomic region) in a number of file formats such as FASTA, NEXUS, MEGA, NBRF, PHYLIP, etc. [from version 1]
- A single MSA storing DNA sequence information (haplotype) along with their frequency, in ARP format (Excoffier and Lischer 2010) [version 6 and newer]
- A multi-MSA File in *.alleles, *.loci and *.fa file formats (generated by PyRAD and STACKS softwares; -Eaton et al. 2014; Catchen et al. 2013), as well as VCF formats (Danecek et al. 2011) [version 6 and newer]

New features in DnaSP in v6 | DnaSP User-Interface | Running DnaSP under Windows, Linux and MacIntosh | Known Bugs in v6 | Source Code |

Go To DnaSP version 5 (version 5.10.1 -March 2009)

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(Abgd, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

Biogeographic hypotheses construction

*(Beast, *Beast, BioGeoBears, Migrate-n)*

Integrative taxonomy

Macroscopy and microscopic data

Literature search

Chemical compounds

DNA polymorphism
(DnaSP)

Phylogenetic relationships
(RAxML, MrBayes, Mega, PhyML)

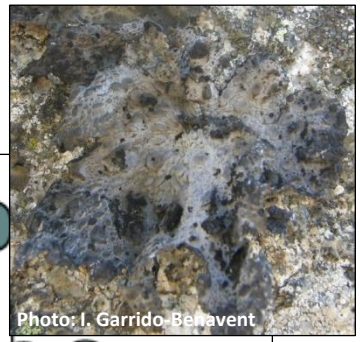
Ecology
(Maxent, Wallace, hypervolume, ecospat, & ENMTools in R)

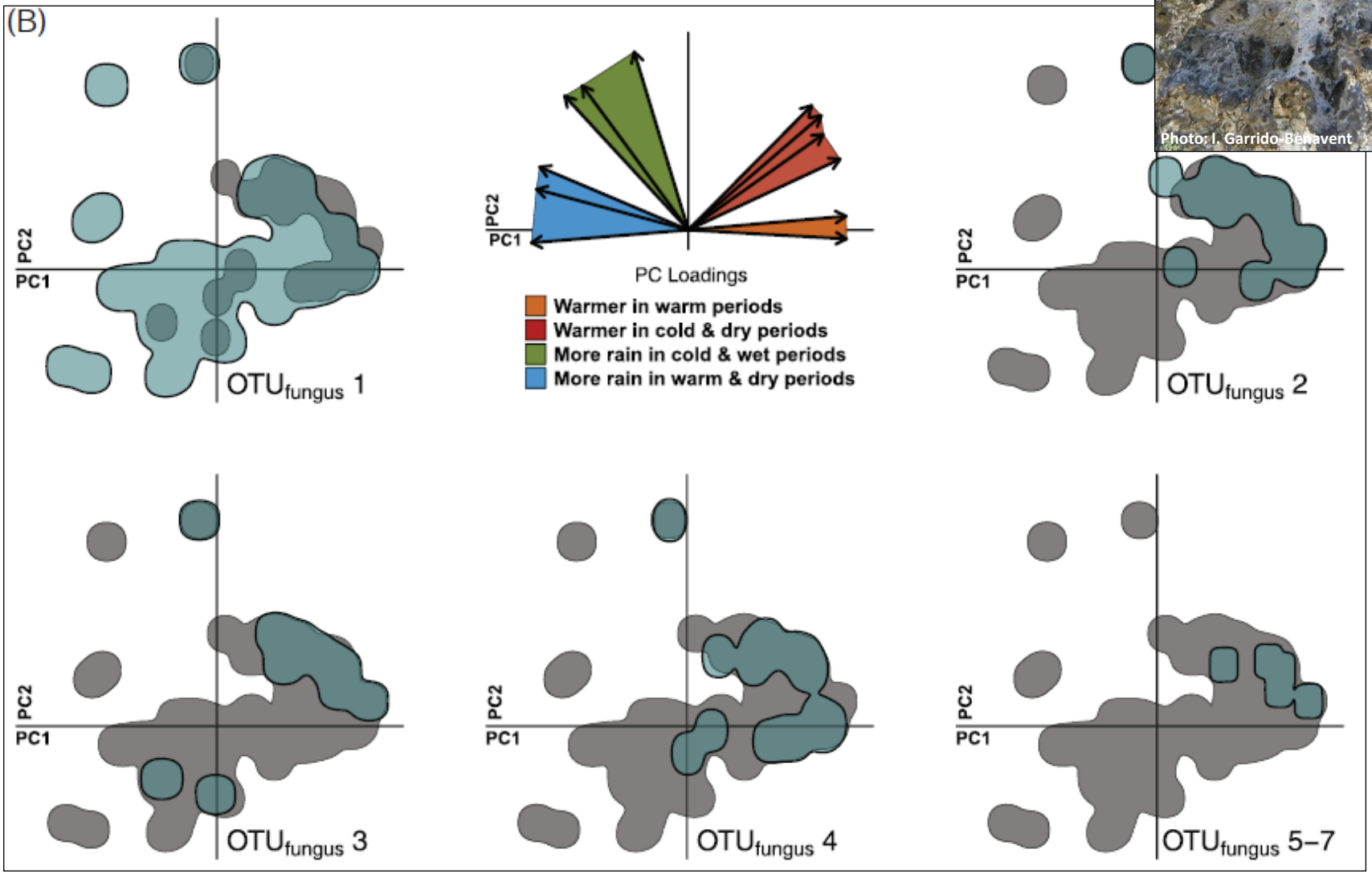
Age estimates
*(Beast, *Beast)*

Recently, a number of tools have been developed to model the ecological niche of species and their spatial distribution. As speciation may have been triggered by different populations of the same species adapting to different ecological conditions without apparent morphological changes, it may be useful to check whether the genetic structure observed in our species correlates well with divergent ecological preferences. In case they correlate, this would further support the distinction of more than one species.

The most popular tool is Maxent, but there are other programs and R functions which should be explored carefully.

Comparison of Hutchinsonian niche hypervolumes

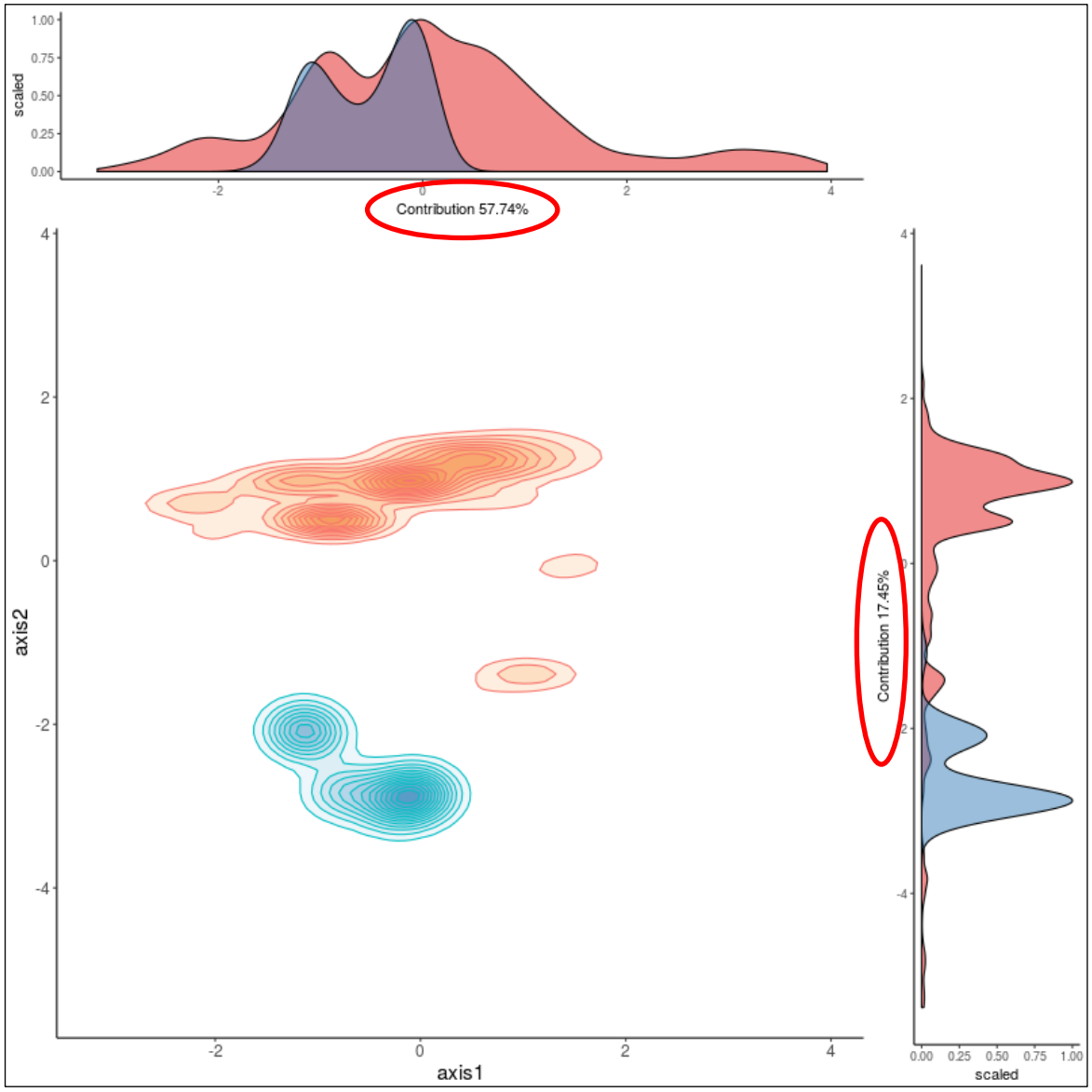




Niche hypervolumes for *L. pustulata* mycobionts (adapted from Rolshausen et al. 2018)

hypervolume R package

Niche overlap between species



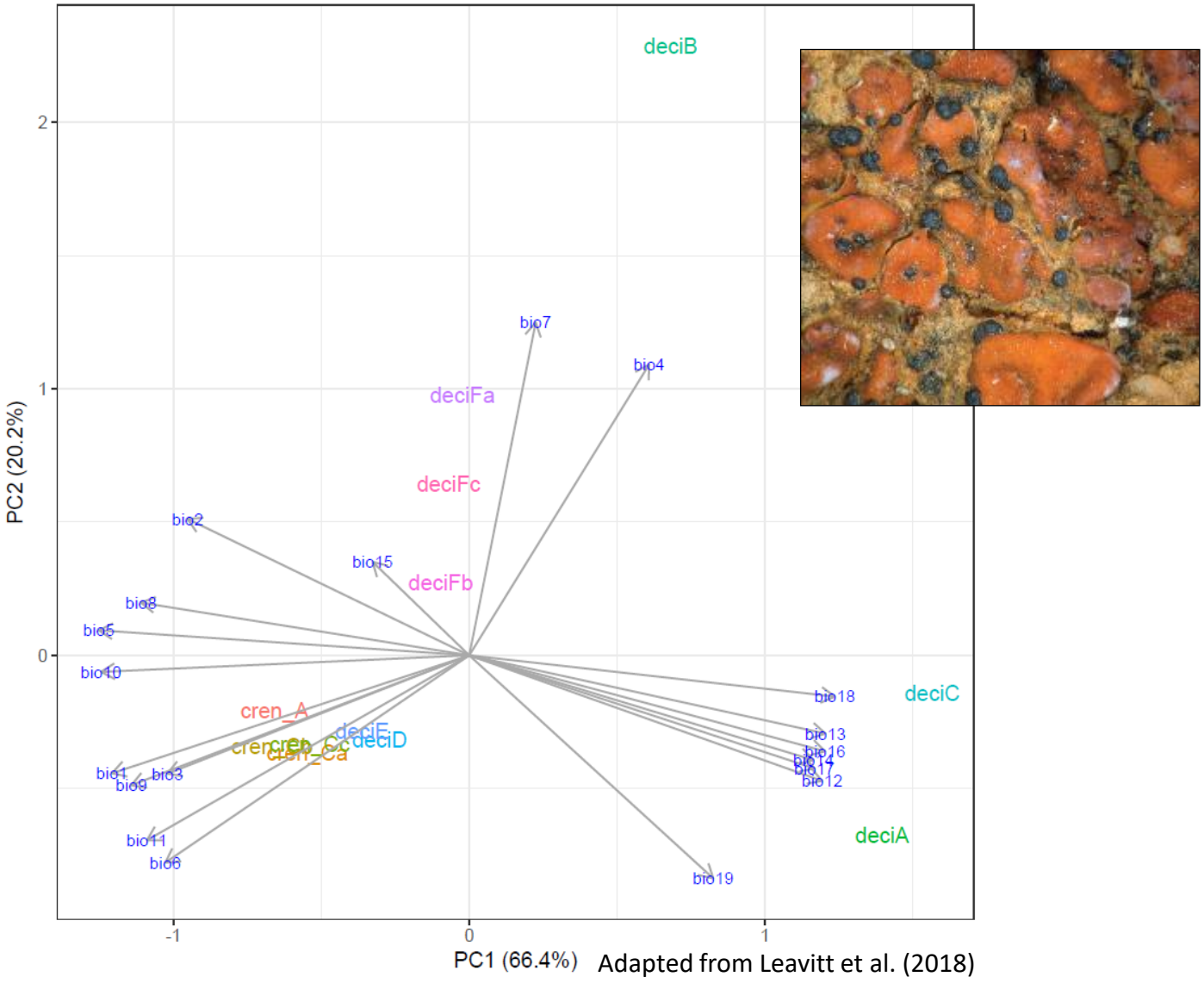
To discuss which environmental variables may have promoted divergence between two (or more) species

Adapted from http://allthiswasfield.blogspot.com/2017/

package *ecospat*

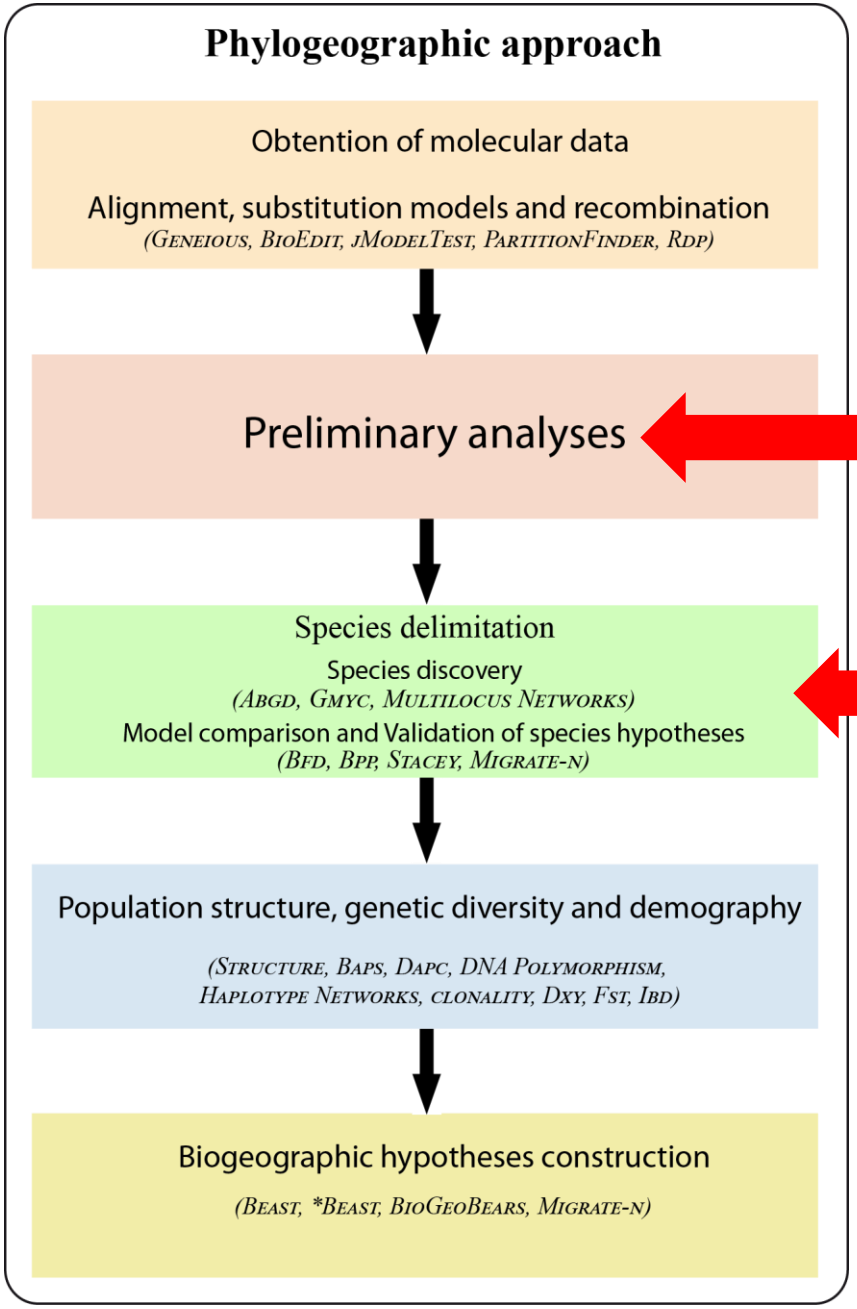
NiceOverPlot function

Phylogenetic PCAs: **visualize the distribution of candidate species or lineages in multidimensional niche space**

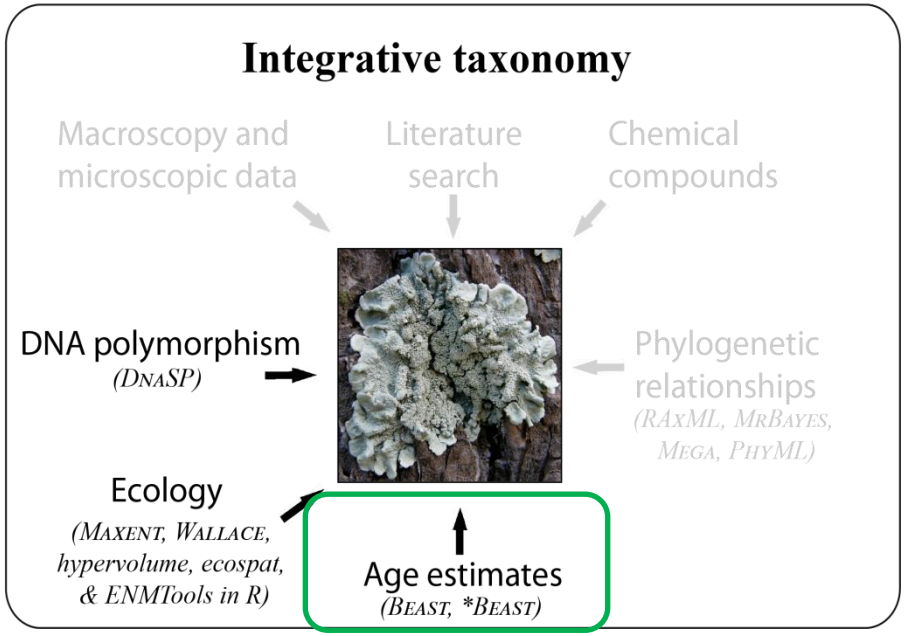


Adapted from Leavitt et al. (2018)

phytools R package



Age estimates are also used to inform species delimitation. This approach has been applied to different groups of fungi.



An approach used in groups of lichenized and non-lichenized fungi
(e.g. Divakar et al. 2017; Hyde et al. 2017)

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(Abgd, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

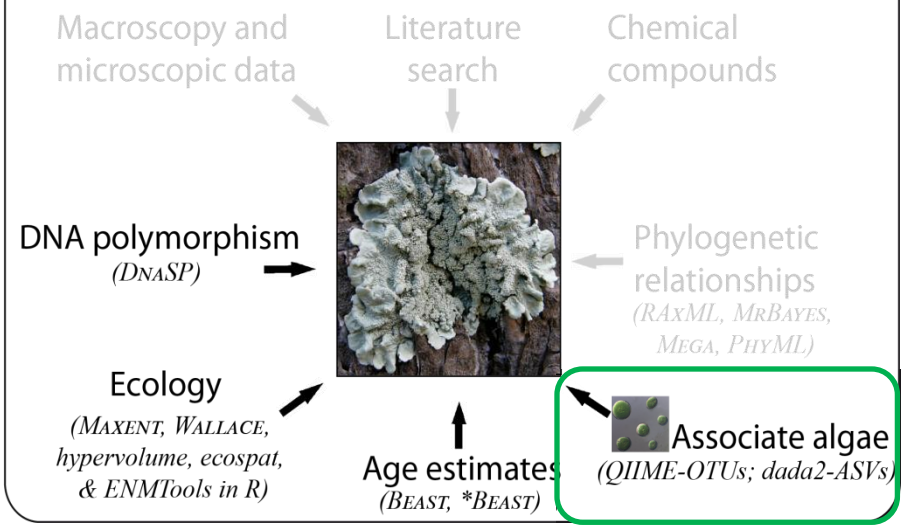
Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

Biogeographic hypotheses construction

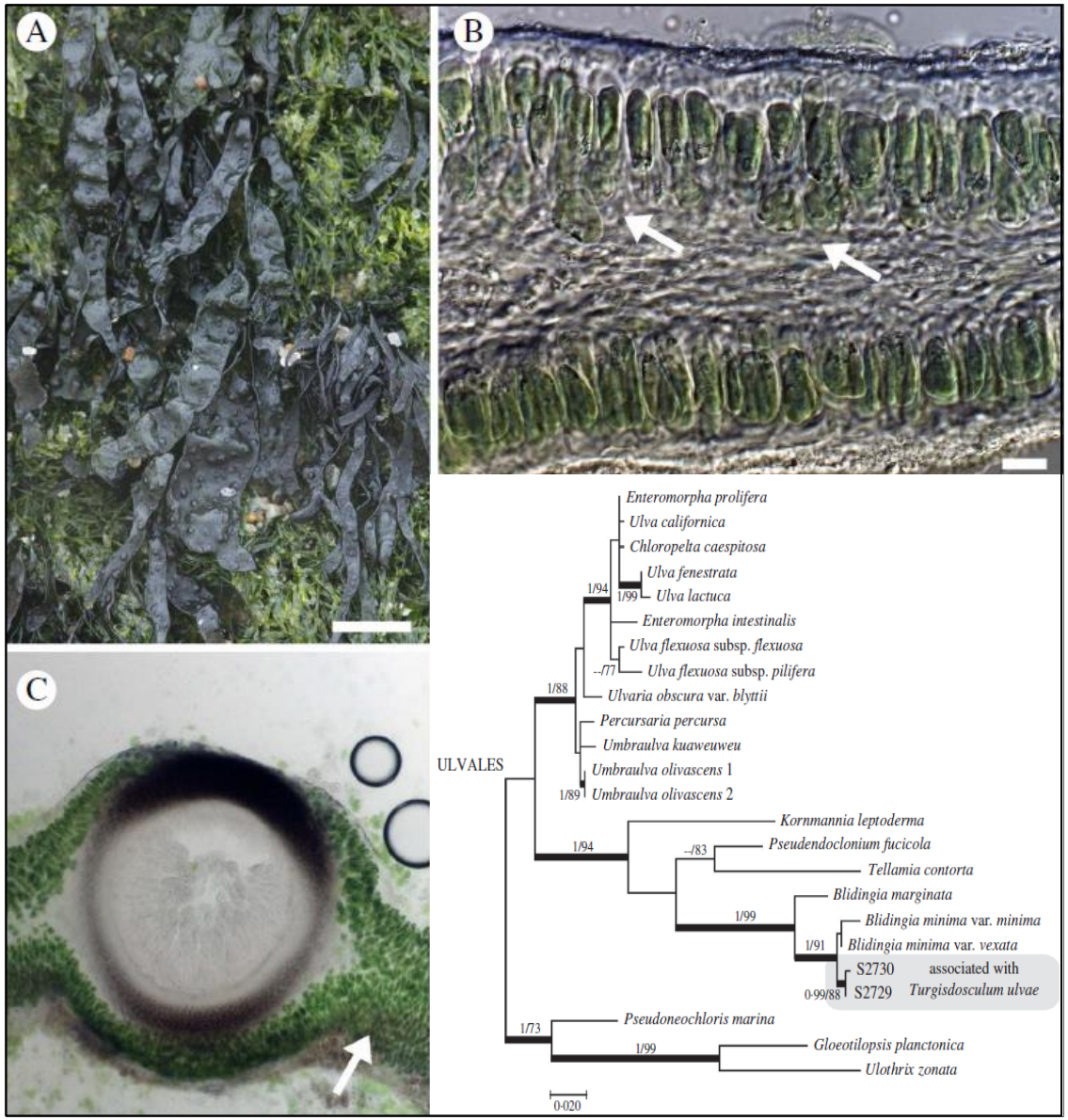
*(Beast, *Beast, BioGeoBears, Migrate-n)*

Integrative taxonomy



A further line of evidence that may help to propose hypotheses on species boundaries in lichenized fungi is the composition of associated algae.

Associated algae inform species delimitation in the fungal partner and viceversa



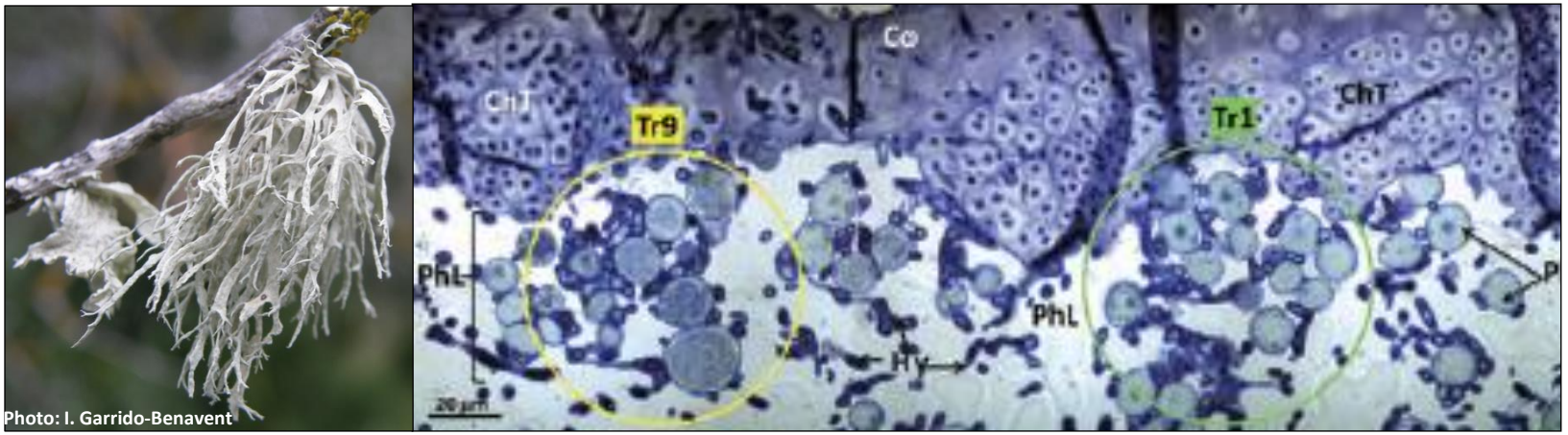
Phylogram based on *rbcL* data and inferred with RAxML (adapted from Pérez-Ortega et al. 2017)

Turgidusculum ulvae
&
Blidingia minima

There are few clear examples where the associated algae clearly point to a specific fungal host because these associations are in principle very strict. This is the case of the *Ulva*-like *Blidingia minima* which associates with the verrucarioid *Turgidusculum ulvae*.

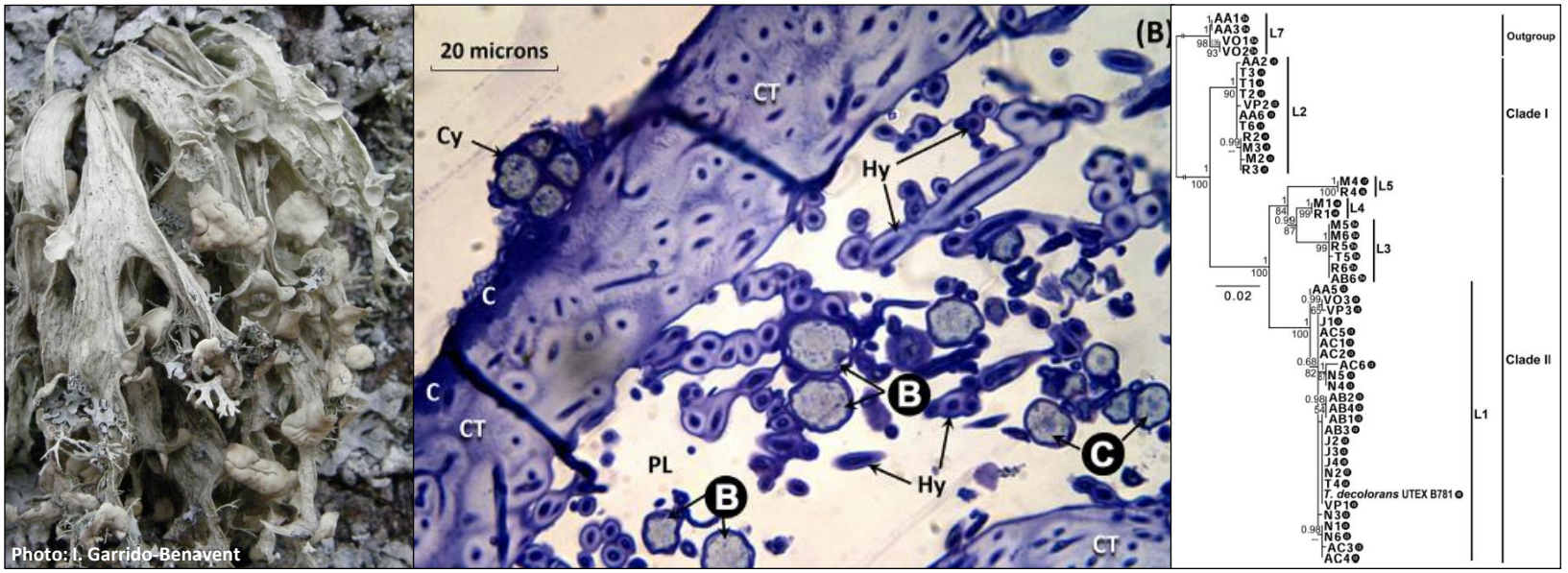
Associated algae inform species delimitation in the fungal partner and viceversa?

However, the usefulness of the taxonomic identity of associated algae for informing species delimitation in fungi becomes more complicated when the associated algae are microscopic and there are thousands of microalgae in the same thallus. In fact, it has been shown that more than one lineage is found in each thallus and that some microalgae are shared between phylogenetically related and unrelated mycobionts.



Ramalina farinacea

Adapted from Casano et al. (2011)

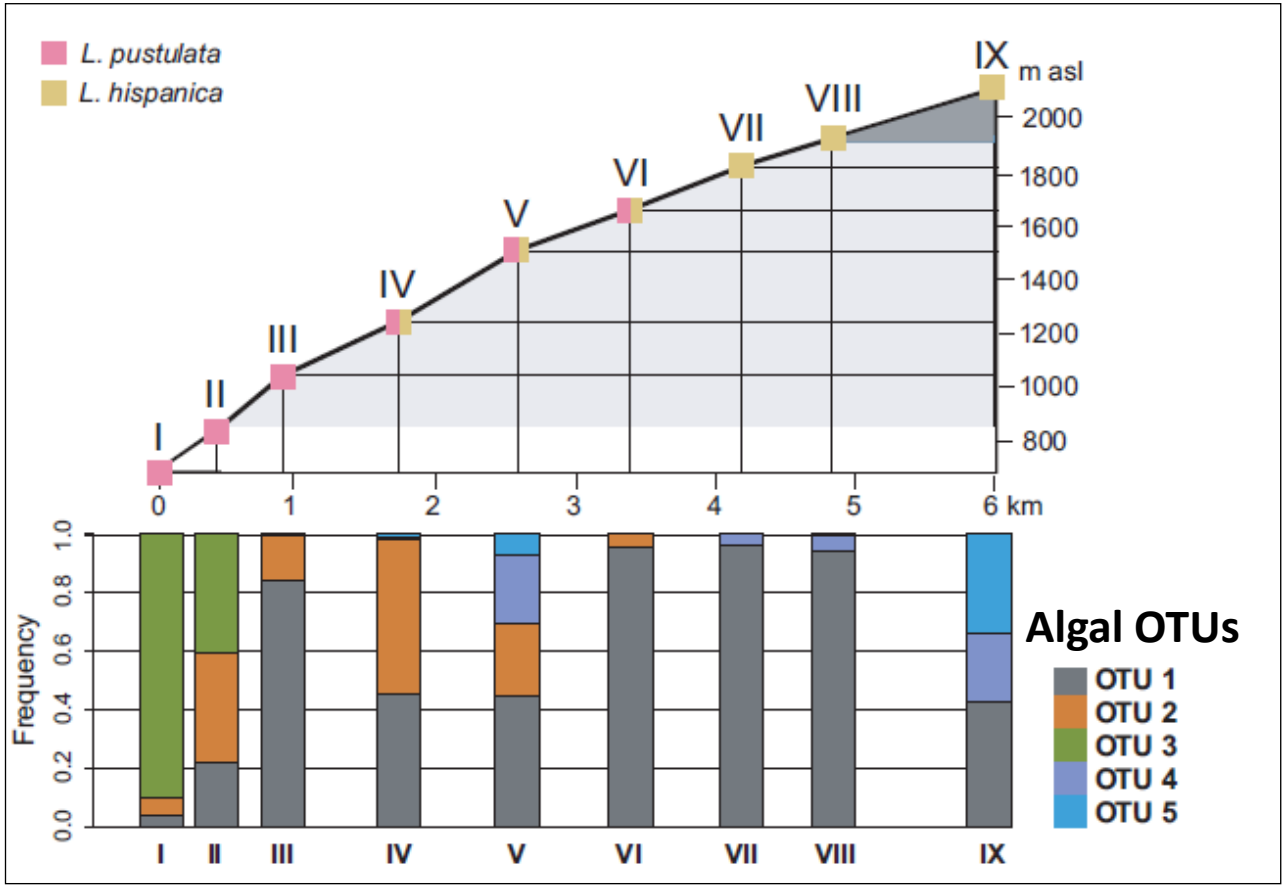


Ramalina fraxinea

Adapted from Català et al. (2016)

The composition of the photobiont community may also characterize certain fungal spp.

In these cases, it is interesting to look at the overall composition of the photobiont community taking advantage of current techniques of DNA metabarcoding. In the following example, although some algal OTUs are shared across species, and that the composition within species varies according to environmental variables, there are still some OTUs that tend to be specific to certain species and also the overall community differs in general between both species.

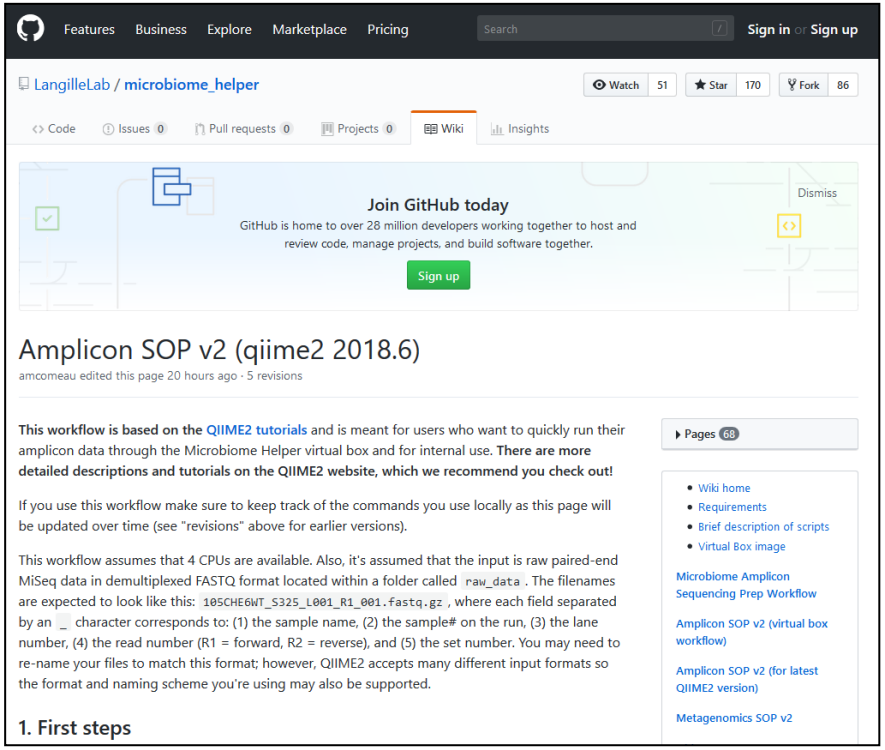


Adapted from Dal Grande et al. (2018)

Algal OTUs inferred from Illumina **DNA metabarcoding data**

Analysis of Illumina metabarcoding data for lichenized fungi and algae

Microbiome Helper is a framework which I found useful when working with DNA metabarcoding data of bacteria, fungi and algae. To install it, you have to download and install first a linux virtual machine, and then install Microbiome Helper on it. The github webpage has tutorials and lots of information. Some time ago Microbiome Helper was devoted to the inference of OTUs with QIIME version 1, but now it offers a more sophisticated way to infer Amplicon Sequence Variants in QIIME2 and dada2 which allow studying metabarcoding data at a finer scale.



Comeau AM, Douglas GM, Langille MGI. 2017. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. mSystems 2(1): e00127-16.

Microbiome Helper

(https://github.com/LangilleLab/microbiome_helper/wiki)

Inference of ASVs

Inference of OTUs

QIIME2 (https://qiime2.org)

dada2 (Callahan et al. 2016)

QIIME1

Oracle Virtual Machine

Linux

Python

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(Abgd, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

Biogeographic hypotheses construction

*(Beast, *Beast, BioGeoBears, Migrate-n)*

Integrative taxonomy

Macroscopy and microscopic data

Literature search

Chemical compounds

DNA polymorphism
(DnaSP)

Phylogenetic relationships
(RAxML, MrBayes, Mega, PhyML)

Ecology
(Maxent, Wallace, hypervolume, ecospat, & ENMTools in R)

Age estimates
*(Beast, *Beast)*

Associate algae
(QIIME-OTUs; dada2-ASVs)
+ microbiota?

Finally, we could even consider the associated microbiota, specially the bacterial communities, as another variable that may inform species delimitation in lichenized fungi. Furthermore, the associated microbiota may also be studied together with the lichenized fungus in phylogeographic studies to explore whether it changes in different geographic regions or under dissimilar ecological conditions.

Phylogeographic approach

Obtention of molecular data

Alignment, substitution models and recombination
(Geneious, BioEdit, jModelTest, PartitionFinder, Rdp)

Preliminary analyses

Species delimitation

Species discovery
(ABGD, Gmyc, Multilocus Networks)

Model comparison and Validation of species hypotheses
(Bfd, Bpp, Stacey, Migrate-n)

Population structure, genetic diversity and demography

(Structure, Baps, Dapc, DNA Polymorphism, Haplotype Networks, clonality, Dxy, Fst, Ibd)

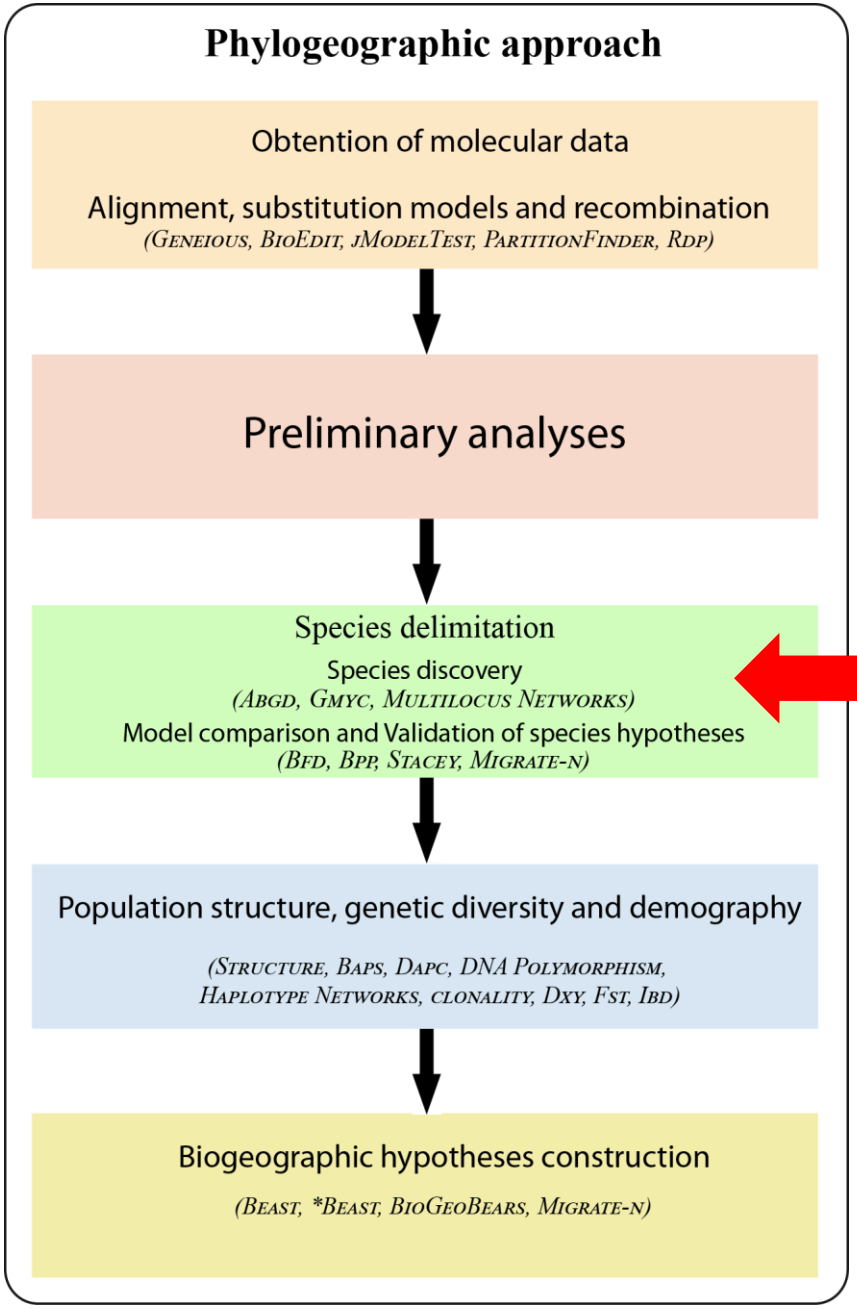
Biogeographic hypotheses construction

*(Beast, *Beast, BioGeoBears, Migrate-n)*

Species discovery

Puillandre et al. (2012)

But we can make an additional use of molecular data using specific software or approaches that infer putative species in our datasets. One of such methods is ABGD, or Automatic Barcode Gap Discovery, which has been widely used to delimit lineages of lichenized fungi and algae.



Other approaches to illustrate the existence of divergent lineages are the construction of multi-locus networks using the software POFAD and SplitsTree.

Species discovery: **multi-locus networks**

POFAD

Phylogeny of Organisms From Allelic Data

© Simon Joly, 2006-2014

Montreal Botanical Garden

Joly & Bruneau (2006)

SplitsTree4

Huson & Bryant (2006)

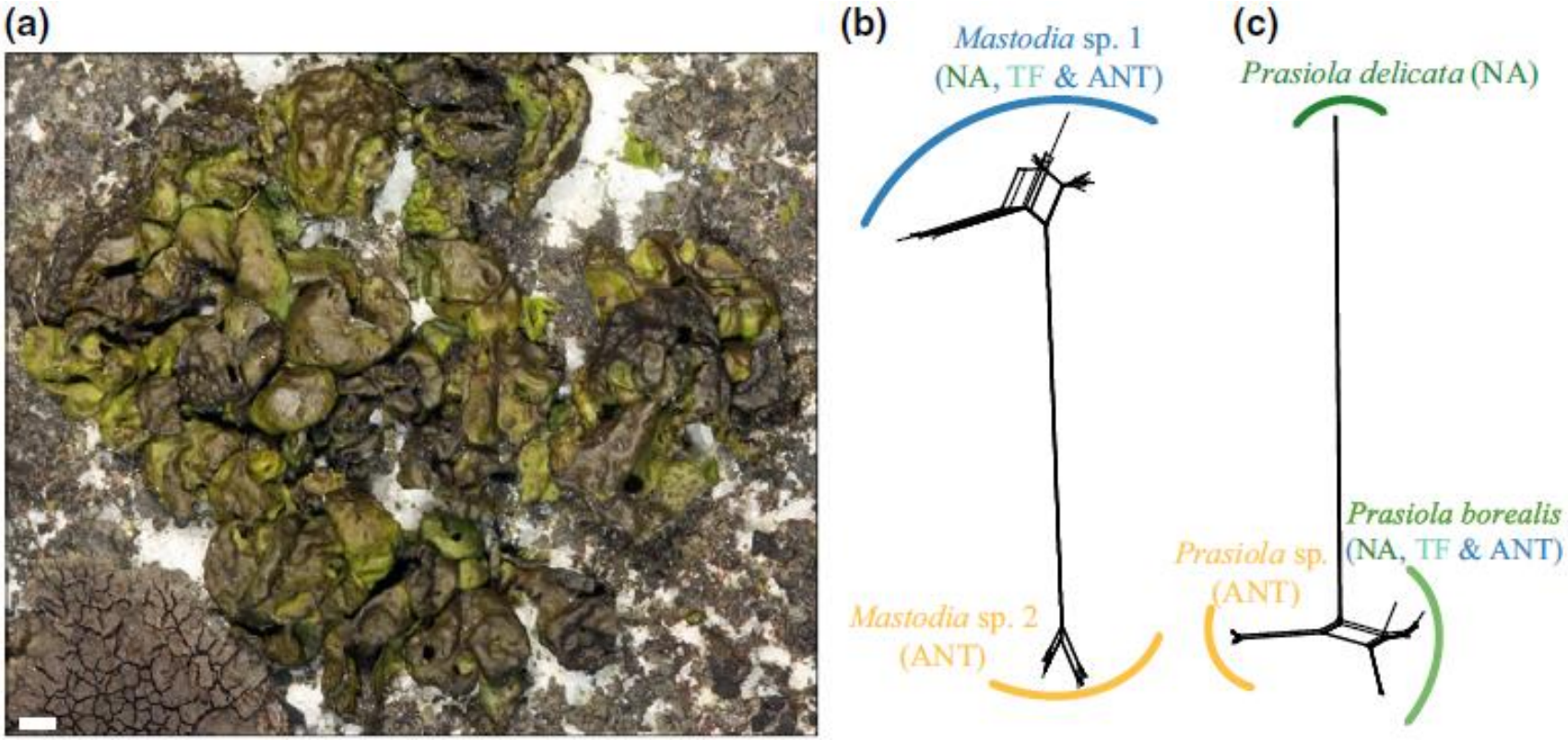
Using multi-locus networks to illustrate species boundaries

POFAD

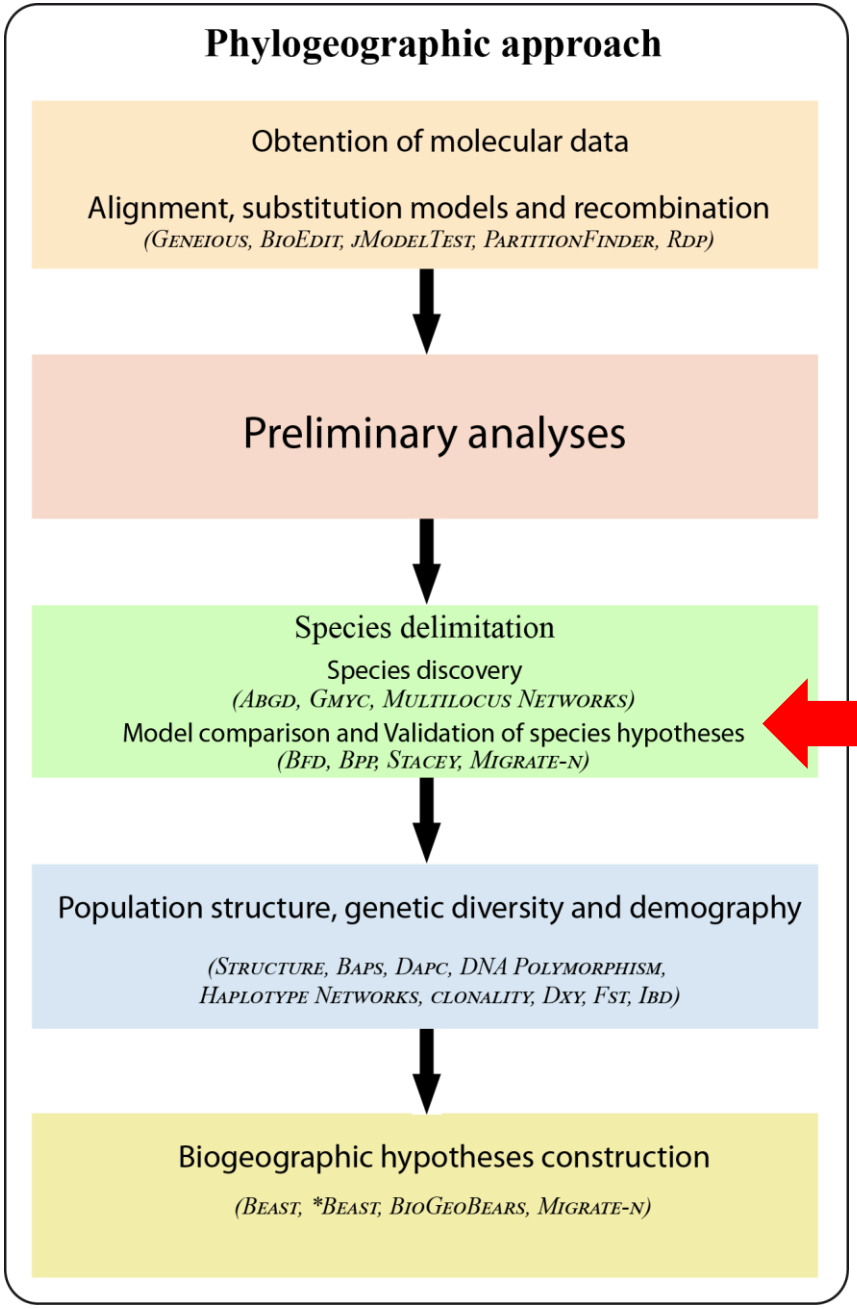
(non)-standardized distance matrices

SplitsTree4

graphical representation



Adapted from Garrido-Benavent et al. (2018)



All of the previous analyses (ABGD, Multi-locus networks) and also the information from the integrative taxonomy steps are used to generate hypotheses that should be tested using multi-locus data and more sophisticated biological and statistical frameworks. For example, using software that operate under the coalescence theory and that may account for incomplete lineage sorting and migration. One of such approaches is the Bayes Factor Delimitation which is implemented in the software StarBEAST.

Species validation:

Bayes Factor Delimitation (BFD)

Grummer et al. (2014)

BEAST

Bayesian Evolutionary Analysis Sampling Trees

StarBEAST: Heled & Drummond (2010)

BFD allows for comparing two or more species delimitation models using BF

TABLE 1 Marginal likelihood and Bayes factor values for two alternative species delimitation hypotheses in the fungal partner of *Mastodia tessellata* and their motivation. Best model highlighted in bold

Model	Distinct species	Motivation	Path Sampling		Stepping-Stone	
			ln (Marginal likelihood)	2ln (Bayes Factor)	ln (Marginal likelihood)	2ln (Bayes Factor)
Model 1 (1 spp.)	n/a	Fungus with a wide distribution	−4455.7	23.6	−4455.7	23.4
Model 2 (2 spp.)	sp1: N. America[a], T. Fuego[b], Antarctica[c] sp2: Antarctica[d]	STRUCTURE multi-locus clustering, ABGD of *nrITS*, multi-locus network	**−4443.9**	n/a	**−4444**	n/a

[a]Individuals with *nrITS* haplotypes: *hap1*, *hap2*, *hap3*, *hap4*, *hap5*.
[b]Individuals with *nrITS* haplotypes: *hap6*, *hap7*, *hap8*, *hap9*, *hap10* and *hap11*.
[c]Individuals with *nrITS* haplotypes: *hap11* and *hap12*.
[d]Individuals with *nrITS* haplotypes: *hap13*, *hap14*, *hap15* and *hap16*.

Adapted from Garrido-Benavent et al. (2018)

Other methods for species validation using the coalescent theory framework

BP&P: Bayesian analysis of genomic sequence data under the multispecies coalescent model

Yang & Rannala (2010, 2014; Rannala & Yang (2013)

- estimation of population size (theta's)
- estimation of species divergence times (tau's)
- species tree estimation

STACEY

Journal of Mathematical Biology

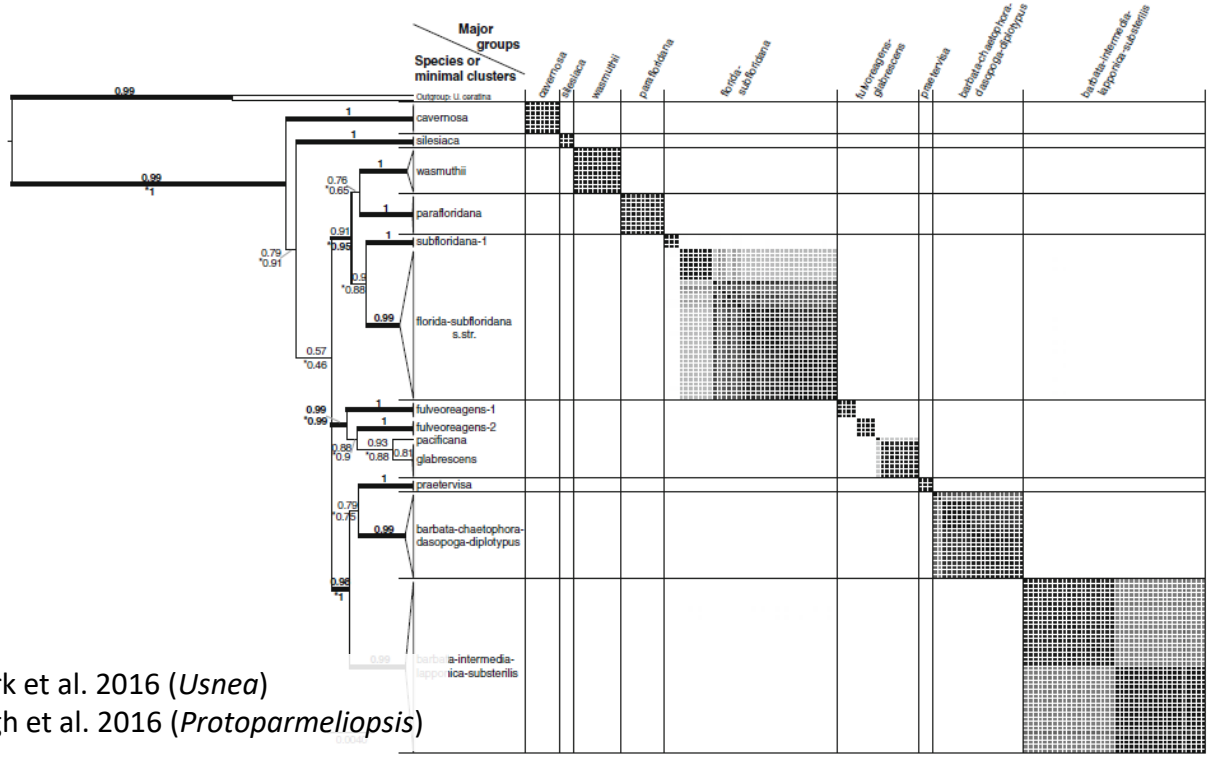
January 2017, Volume 74, Issue 1–2, pp 447–467 | Cite as

Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent

Authors | Authors and affiliations

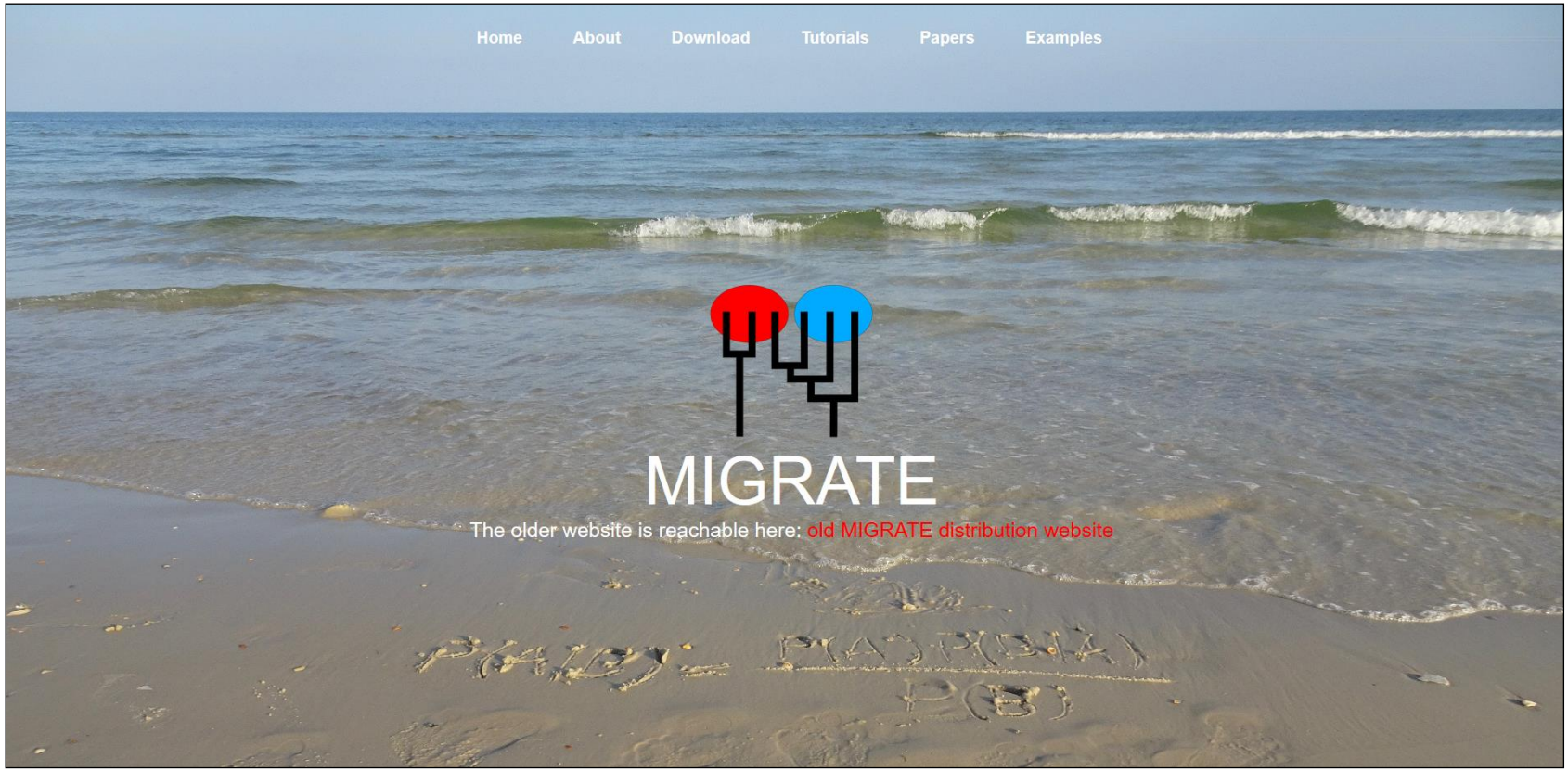
Graham Jones

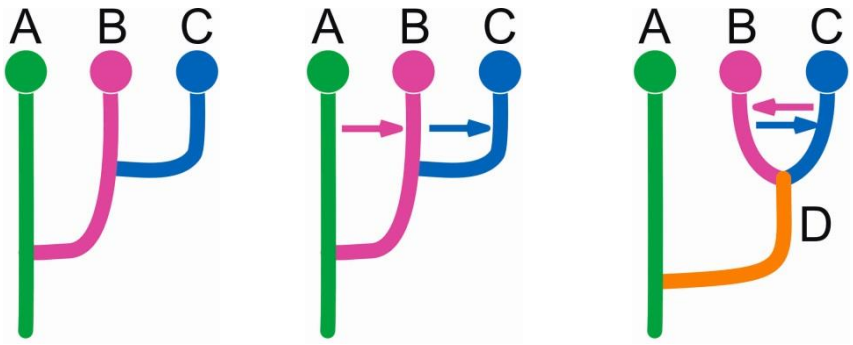
Jones (2017)



Mark et al. 2016 (*Usnea*)
Singh et al. 2016 (*Protoparmeliopsis*)

Migrate-n is useful to test models of species divergence as well

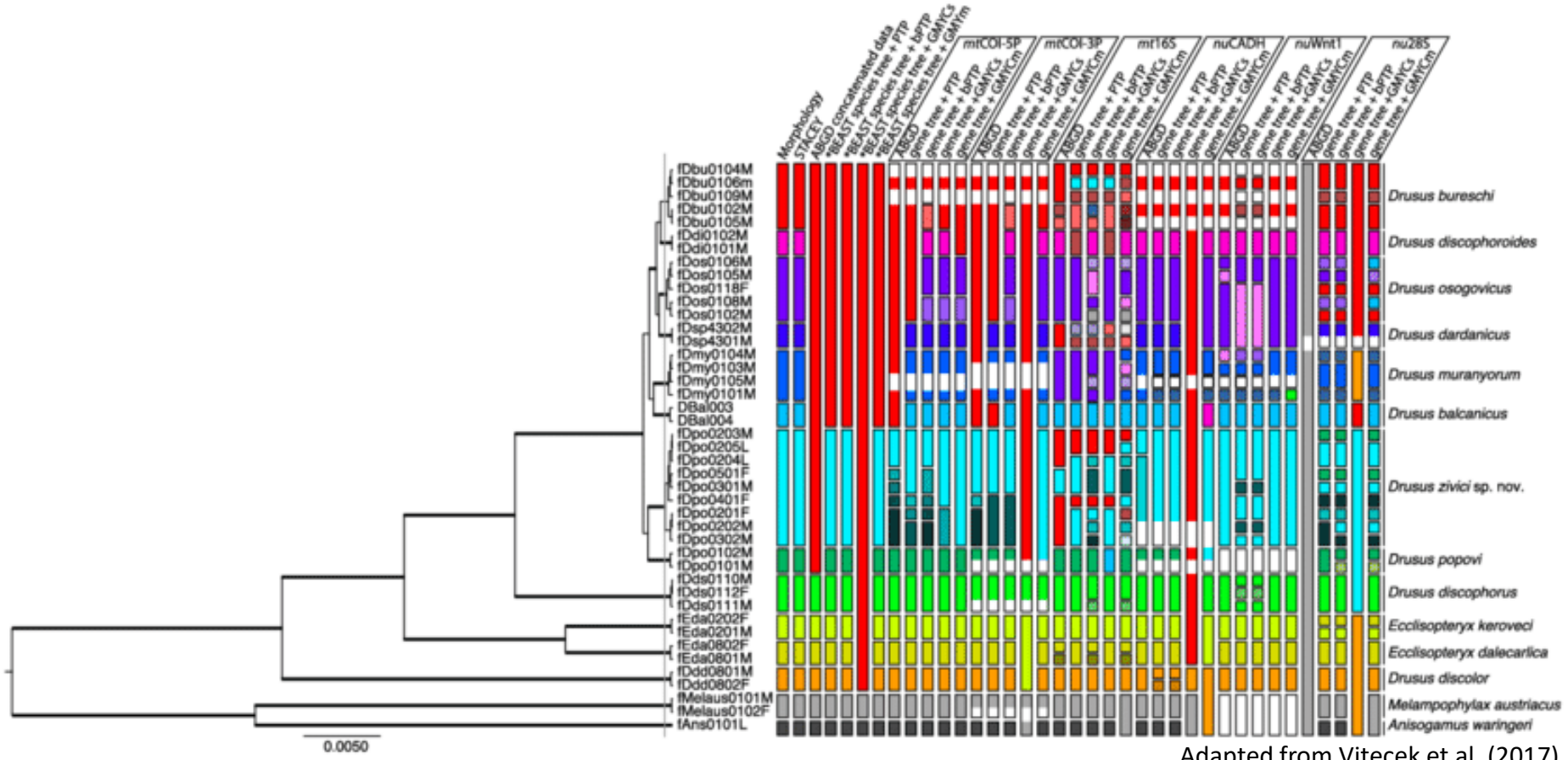




Version 4.x allows for specifying models with divergence

Beerli (2006); Beerli & Palczewski (2010)

How many analyses do we carry out???

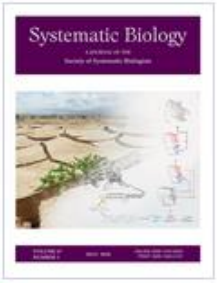


Adapted from Vitecek et al. (2017)

- Species discovery: ABGD (distance), data from integrative taxonomy
- Species validation: BFD, BP&P OR MIGRATE-N

Estimating phylogenies or species networks under ILS and gene flow

Several methods have been recently developed to estimate phylogenies and species networks under Incomplete Lineage Sorting and gene flow. The analyses are applicable to multiple species and/or multiple populations per species.



Volume 67, Issue 3
May 2018

Coestimating Reticulate Phylogenies and Gene Trees from Multilocus Sequence Data

FREE

Dingqiao Wen, Luay Nakhleh ✉

Inferring Phylogenetic Networks Using PhyloNet

Dingqiao Wen, Yun Yu, Jiafan Zhu, Luay Nakhleh ✉

Volume 67, Issue 4
July 2018

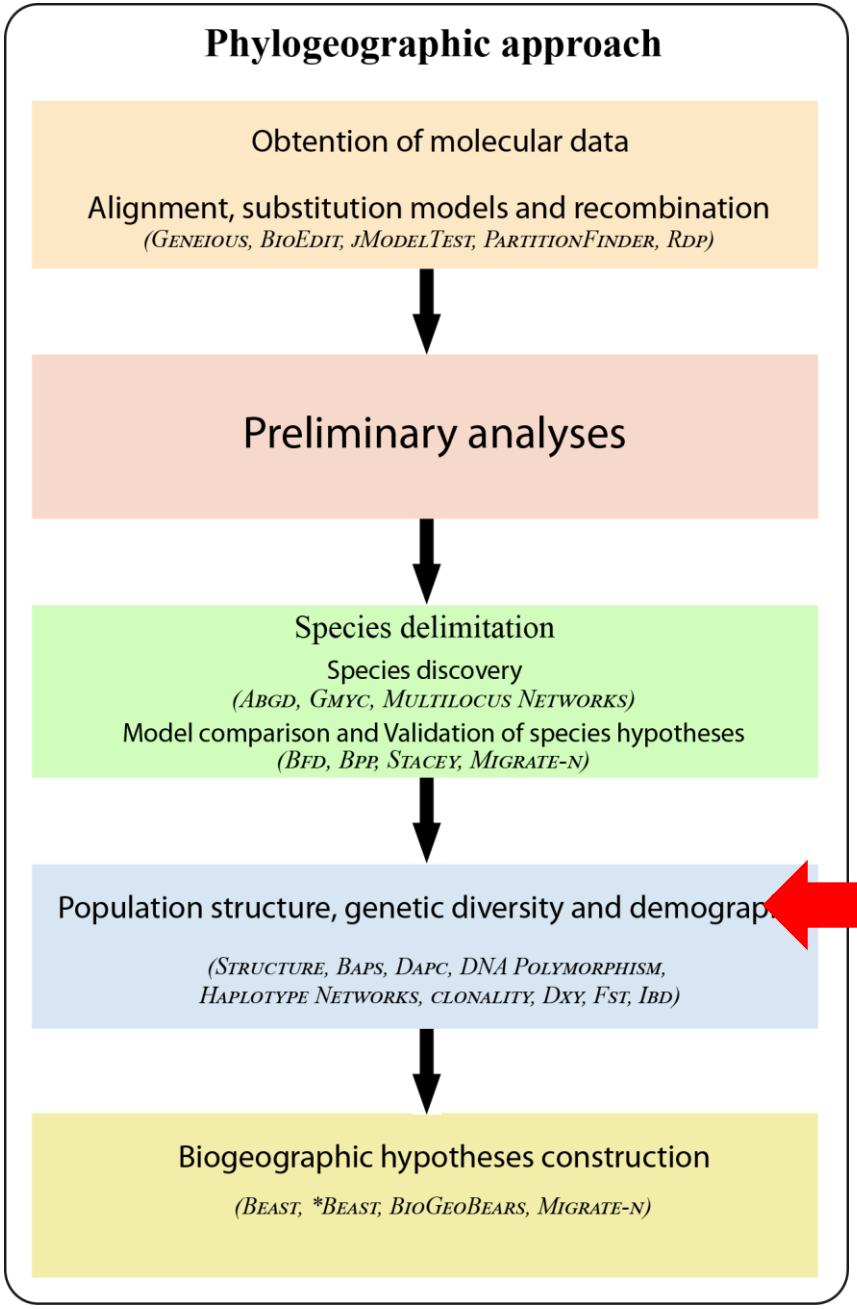
Divergence Estimation in the Presence of Incomplete Lineage Sorting and Migration

Graham R Jones ✉

Volume 35, Issue 11
November 2018

Phylogeny Estimation by Integration over Isolation with Migration Models

Jody Hey ✉, Yujin Chung, Arun Sethuraman, Joseph Lachance, Sarah Tishkoff, Vitor C Sousa, Yong Wang



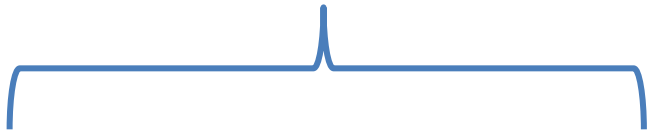
After having carefully assessed species boundaries in our phylogeographic dataset, the next step is more related with population genetics. Here, we conduct a series of analyses devoted to characterize population structure, genetic diversity in general and for each geographic region, and to detect demographic changes through time.

Population assignment and structure

Alternative methods for population assignment and structure

Based on pre-defined population genetic models
(generally assume unlinked markers and panmictic populations)

Based on distance (model-free)
(without these assumptions)



Structure Software

Pritchard Lab
(Stanford University)

(ad-)mixture model based on a multi-locus matrix of **haplotype numbers**

BAPS Software

Bayesian Statistics Group
(University of Helsinki)

(ad-)mixture models based on single or multi-locus matrices of **SNP data**

DAPC

Jombart et al. (2010, 2015)

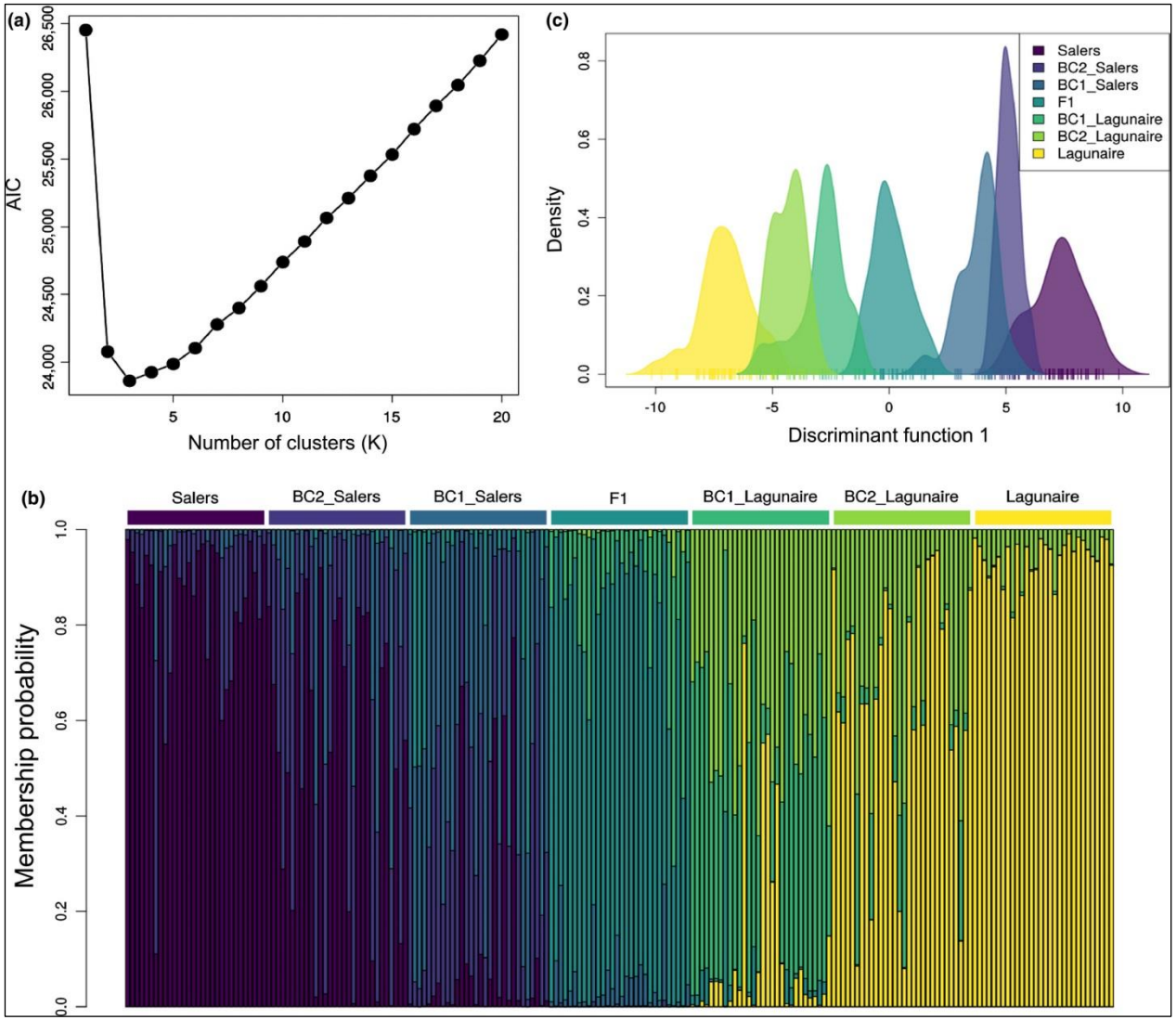
Identification of clusters using multi-locus matrices of **haplotype numbers**

Bayesian & computationally-intensive

Computationally much less intensive

Huelsenbeck, John P., Peter Andolfatto, and Edna T. Huelsenbeck. "**Structurama**: Bayesian inference of population structure." *Evolutionary Bioinformatics* 7 (2011): EBO-S6761.

snapclust: a new approach combining model- and distance-based methods

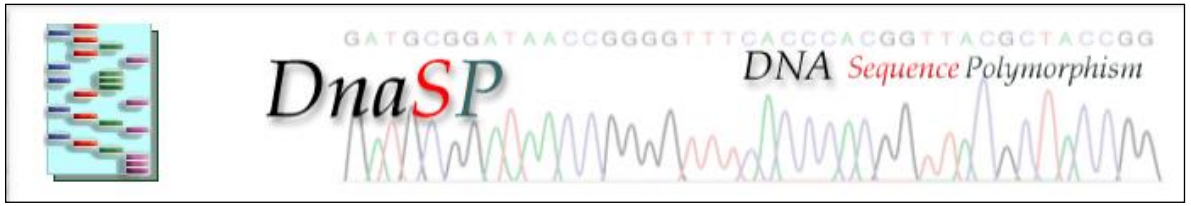


Adapted from Beugin et al. (2018)

-SNP
-Microsatellites
-AA sequences

package *adegenet*

DNA polymorphism



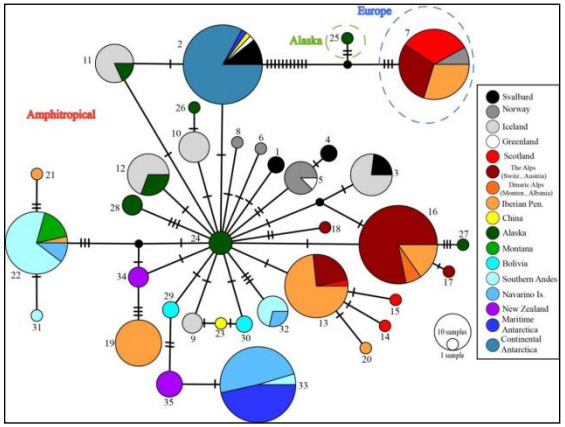
Genetic diversity indices:

- # of segregating sites (s)
- # of haplotypes (h)
- Haplotype diversity (Hd)
- Average number of nucleotide differences (k)
- Nucleotide diversity (π)
- # of parsimony informative sites

Haplotype networks

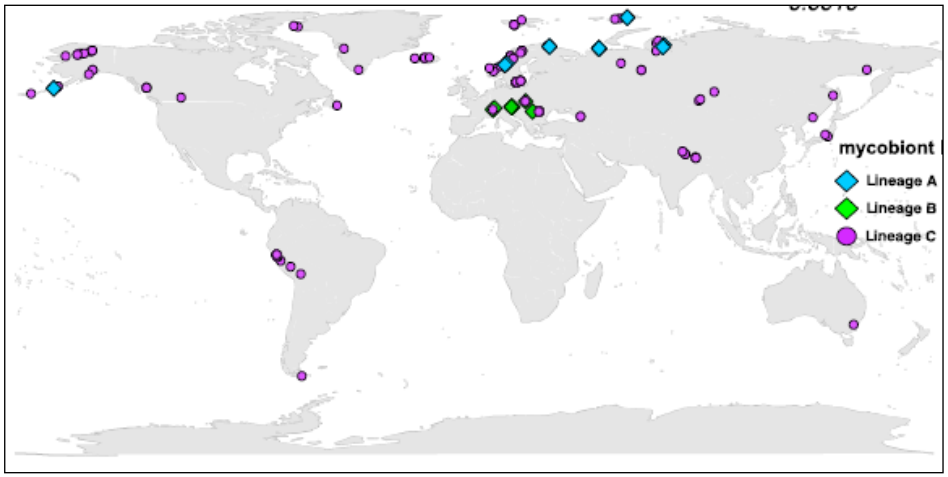
Single-locus haplotype dataset

TCS (Clement et al. 2002)
PopART v.1.7
(Leigh & Bryant 2015)



Estimating clonal reproduction

I_A method (Avise & Wollenberg 1997)



Sequence data Adapted from Onuţ-Brännström et al. (2017)

rBarD (unbiased index of association, Agapow & Burt 2001)

Microsatellites data Adapted from Alors et al. (2017)

package *poppr*

Quantifying genetic divergence and differentiation

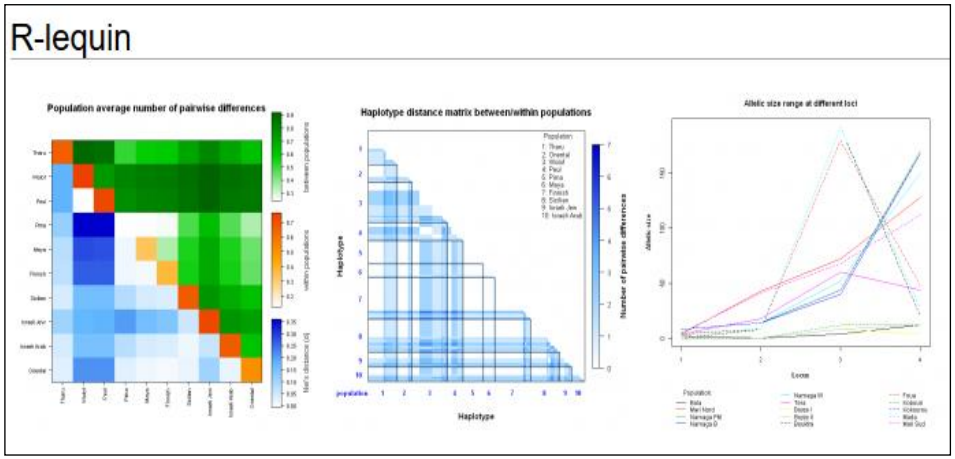
Arlequin (Excoffier &Lischer 2010)

Genetic divergence

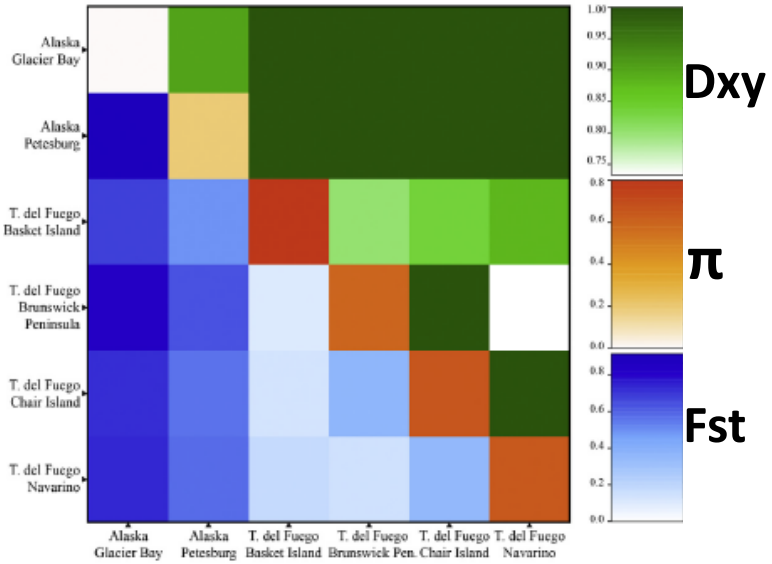
Dxy (Nei 1987): average number of nucleotide substitutions per site between sampling localities

Genetic differentiation

Fst (Weir & Cockerham 1984): estimator H for Wright's fixation index based on allele frequencies

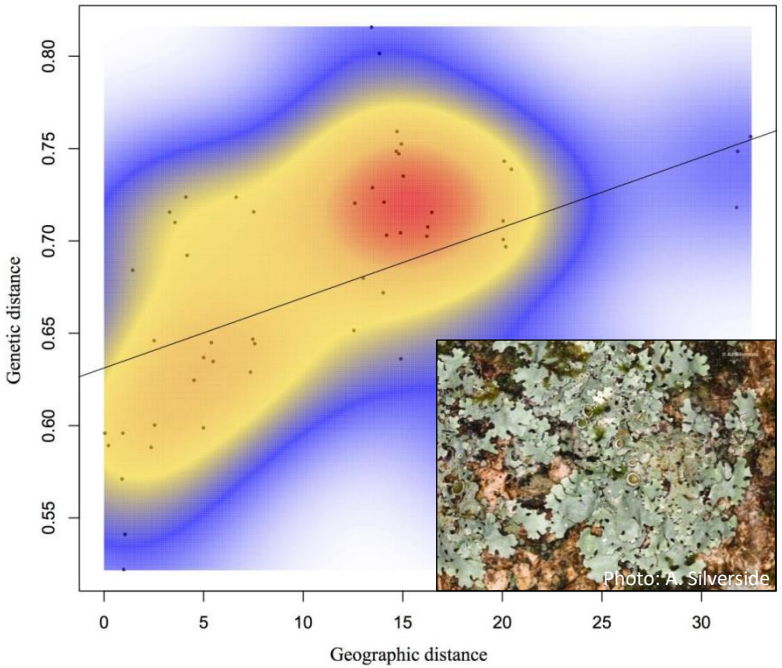


***R-lequin*: collection of R functions to parse Arlequin output files and produce high quality graphics** (http://heidi.chnebu.ch/doku.php?id=r-lequin)



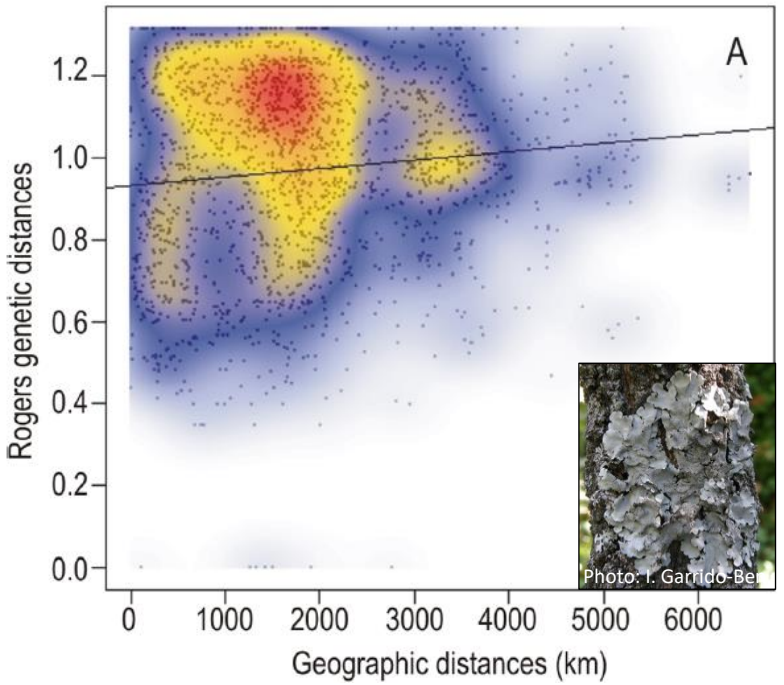
Testing for Isolation by Distance (IBD)

Parmelina carporrhizans and *P. tiliacea* in the Mediterranean and Macaronesian regions



- **Clear** IBD pattern (r = 0.472, *P* = 0.005)
- Single cloud of points indicates a **continuous cline** of genetic differentiation

(Adapted from Alors et al. 2017)



- **Weak** IBD pattern (r = 0.111, p = 0.003)
- Incipient patchy pattern of points indicates the existence of distant and differentiated populations (Adapted from Núñez-Zapata et al. 2015)

mantel.randtest (Mantel test, package *adegenet* in R)

Measuring local densities of distances (function *kde2d*) and plotting in function *image* in the R package *MASS*)

Take also a look at the R package *MEMGENE* (Galpern et al. 2014) to describe the geographic distribution of genetic variability at landscape scale (see Rolshausen et al. 2018 for a example in lichenized algae)

To partial out the relative contribution of climate and geography on the variation of genetic structure

Full and partial **redundancy analyses** (**RDA**, Legendre & Fortin 2010)

Lobaria pulmonaria in the Iberian Peninsula

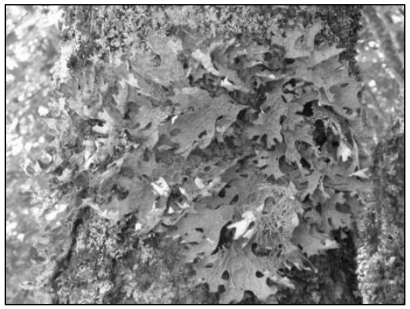
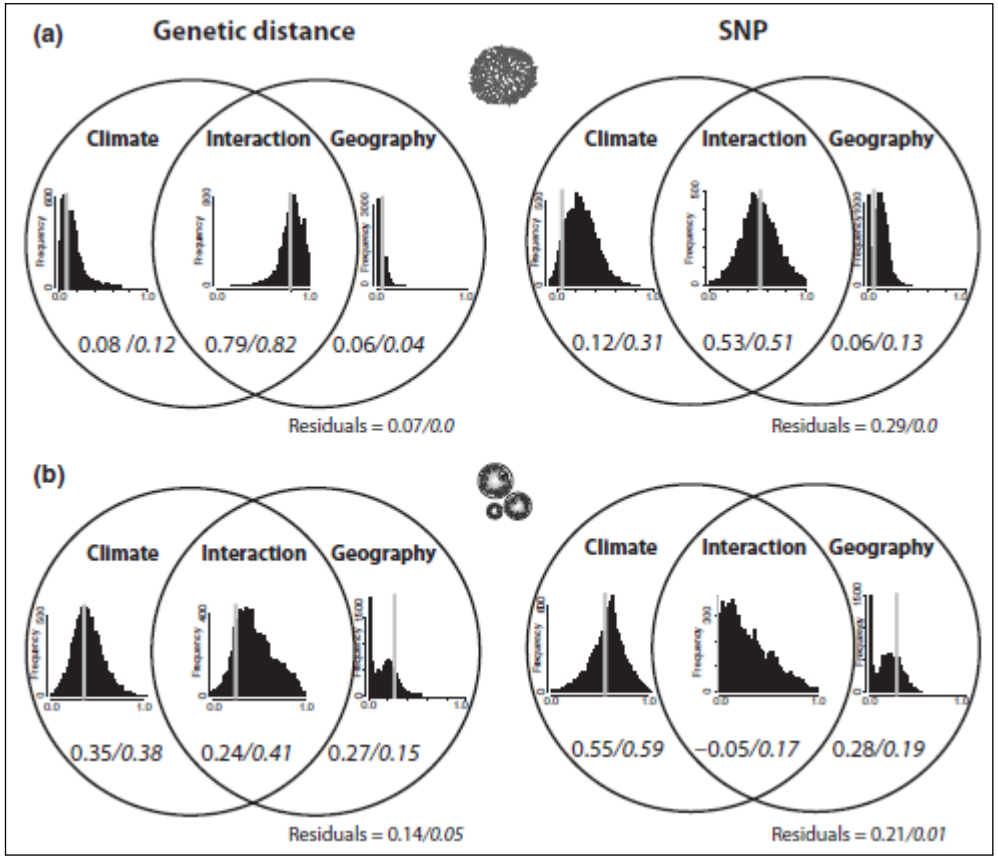


Table 5 – Explained variance by each RDA model.

	Variance	Probability
Total explainable (RDA1)	43.6	<0.01
Pure environment (RDA2)	26.4	<0.05
Pure geography (RDA3)	7.6	<0.1
Joint climate/geography (RDA1-RDA2-RDA3)	9.6	

Adapted from Otálora et al. (2015)



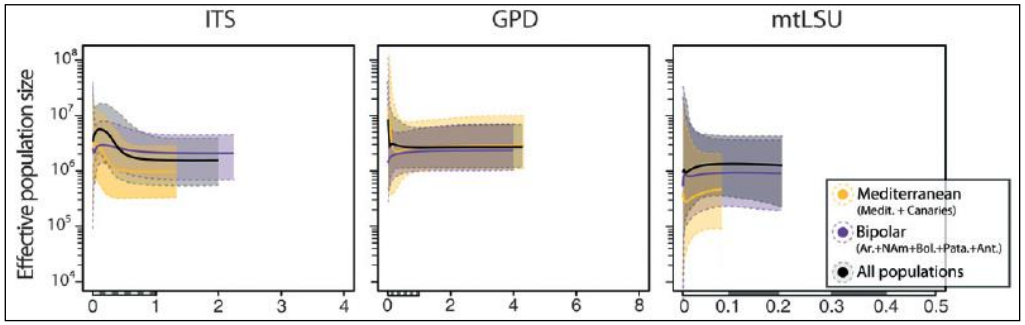
Adapted from Fernández-Mendoza et al. (2011)

varpart, *rda*, *anova.CCA* functions in the R package *vegan*

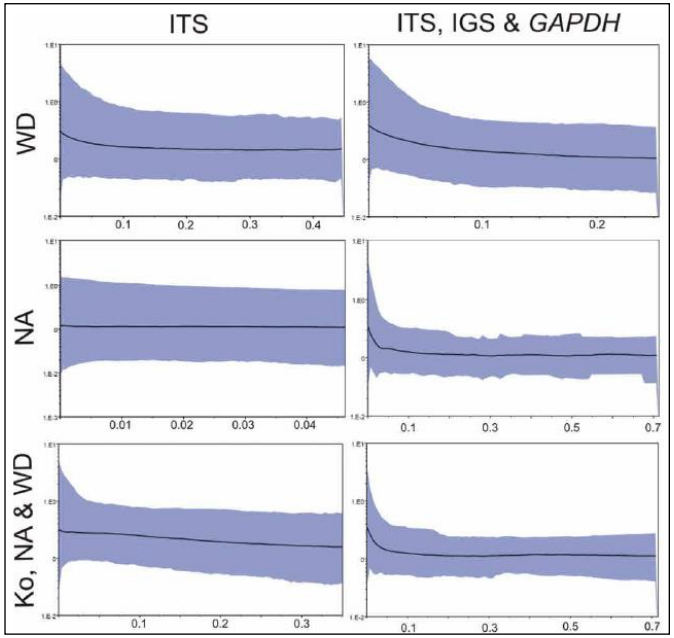
Demography and deviations from neutrality

Bayesian Skyline Plots

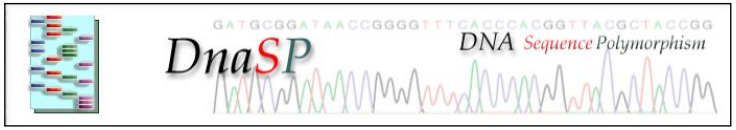
BEAST
Bayesian Evolutionary Analysis Sampling Trees



Adapted from Fernández-Mendoza & Printzen (2013)



Adapted from Boluda et al. (2018)

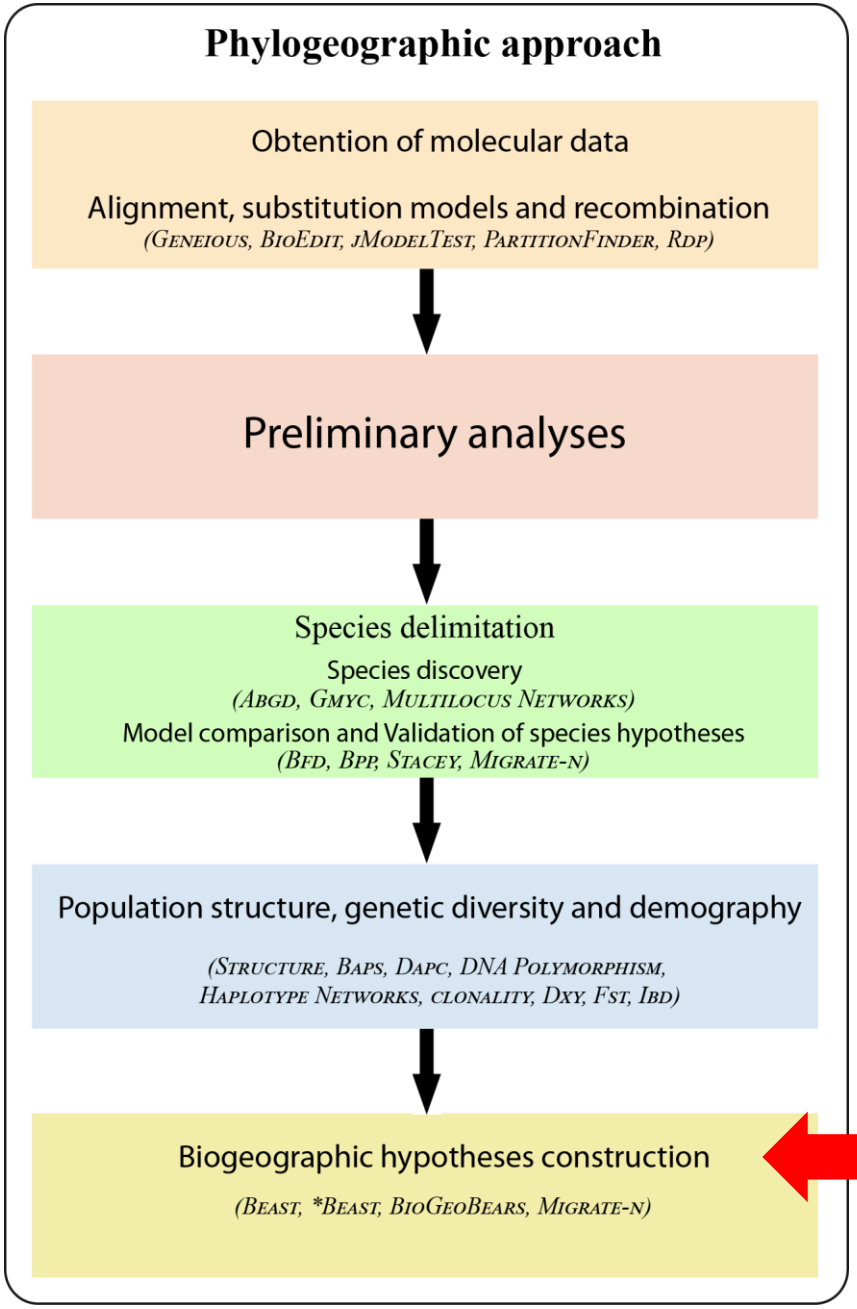


Tajima's D and **Fu's Fs** statistics.

- number of segregating sites
- significance based on 10^x coalescent simulations

\+ **D and Fs values:** diversifying selection or a recent bottleneck

\- **D and Fs values:** purifying selection or a recent expansion



And finally, there is a set of analyses to build and test biogeographic hypotheses *per se*. These are dating, the reconstruction of ancestral ranges, and the inference of migration.

Dating, ancestral range reconstruction and migration

Dating

BEAST

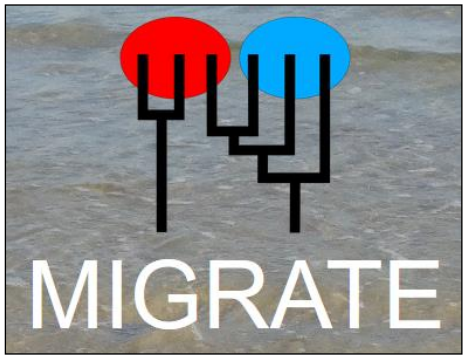
Bayesian Evolutionary Analysis Sampling Trees

BEAST v. 1.8: Drummond et al. (2012)
StarBEAST: Heled & Drummond (2010)

Ancestral range reconstruction

BioGeoBEARS (Matzke 2013)

Migration



Beerli (2006);
Beerli & Palczewski (2010)

Dating

BEAST v. 1.8: Drummond et al. (2012)
StarBEAST: Heled & Drummond (2010)

Markers

1 locus

vs

≥ 2 loci

Type of calibration

Primary calibration
(fossil data)

vs

Secondary calibration
(age estimate
or substitution rate)

Dating

BEAST v. 1.8: Drumond et al. (2012)

StarBEAST: Heled & Drummond (2010)

Test for a strict molecular clock for EACH locus

Table 2 Test for a strict molecular clock for each locus in Dataset A and B conducted in MEGA 5.0

	ML estimate				MrBayes consensus			
nrITS GTR+Γ+I (Dataset A)	ln*L*	Param	(+Γ)	(+I)	ln*L*	Param	(+Γ)	(+I)
With clock	−7095.856	79	1.115	0.36	−7099.121	79	1.191	0.35
Without clock	−6960.321	147	1.15	0.32	−6960.203	147	1.2	0.33
P (Ho: = rates)	$5.2e^{-11}$*				$8.96e^{-12}$*			
nrITS TN93+Γ+I (Dataset B)	ln*L*	Param	(+Γ)	(+I)	ln*L*	Param	(+Γ)	(+I)
With Clock	−5444.788	40	0.977	0.28	−5452.139	40	0.913	0.32
Without clock	−5388.285	75	0.96	0.32	−5363.874	75	0.89	0.30
P (Ho: = rates)	$8.65e^{-4}$*				$3.63e^{-11}$*			
nuLSU GTR+Γ+I (Dataset B)	ln*L*	Param	(+Γ)	(+I)	ln*L*	Param	(+Γ)	(+I)
With clock	−2753.380	42	0.778	0.63	−2746.279	42	0.793	0.64
Without clock	−2722.074	73	0.76	0.64	−2718.789	73	0.78	0.64
P (Ho: = rates)	0.45				0.72			
mtSSU HKY+Γ+I (Dataset B)	ln*L*	Param	(+Γ)	(+I)	ln*L*	Param	(+Γ)	(+I)
With clock	−3006.041	42	0.706	0.59	−3025.628	42	0.789	0.63
Without clock	−2948.207	77	0.81	0.64	−2946.274	77	0.82	0.64
P (Ho: = rates)	$4.9e^{-4}$*				$7.73e^{-9}$*			

Tested under two different topologies (ML and Bayesian). *denotes rejection of the null hypothesis (i.e., equal rates)

Adapted from Garrido-Benavent et al. (2016)

Strict vs uncorrelated relaxed lognormal molecular clock

Prior settings There is no general rule. Literature review.

Run settings Depending on the amount and complexity of data, run at least one analysis with chains 1–5×10^7generations long

Dating

BEAST v. 1.8: Drummond et al. (2012)
StarBEAST: Heled & Drummond (2010)

Run the analysis/-es

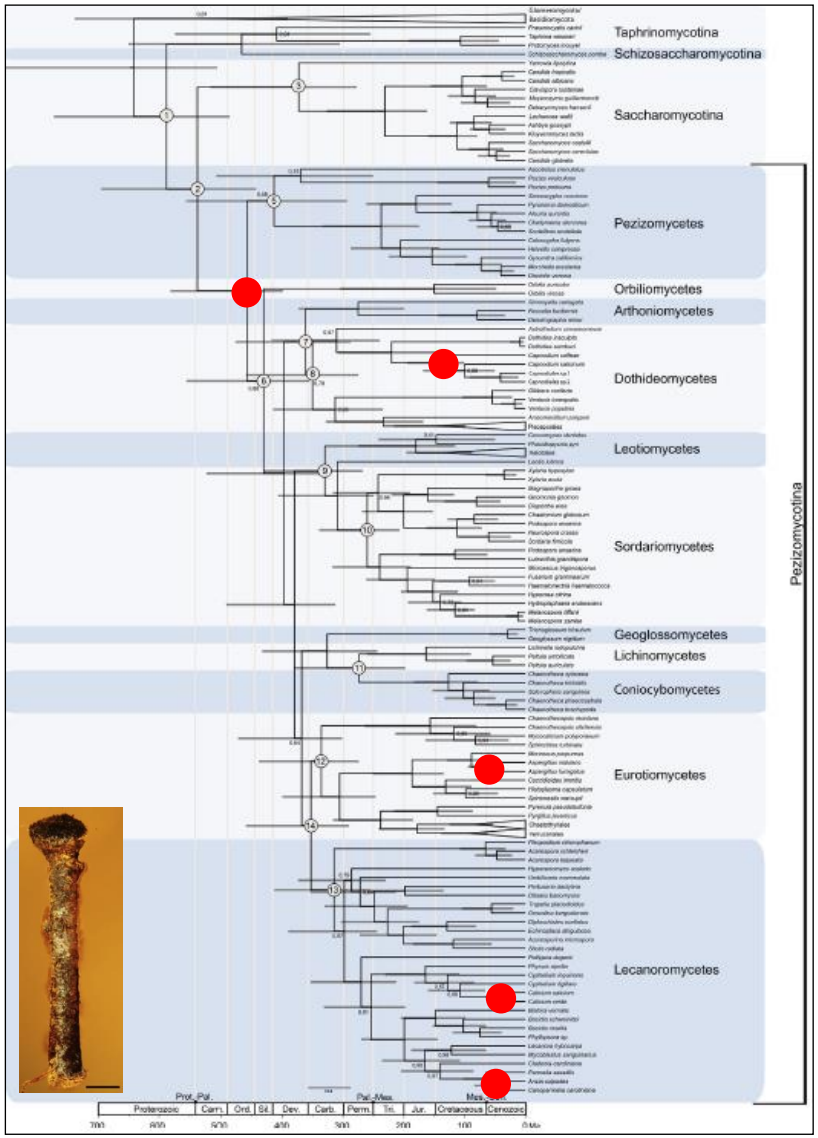
BEAUti → XML file →

Check for convergence of chains

The effective sample sizes (**ESS**) of each parameter must be at least 200

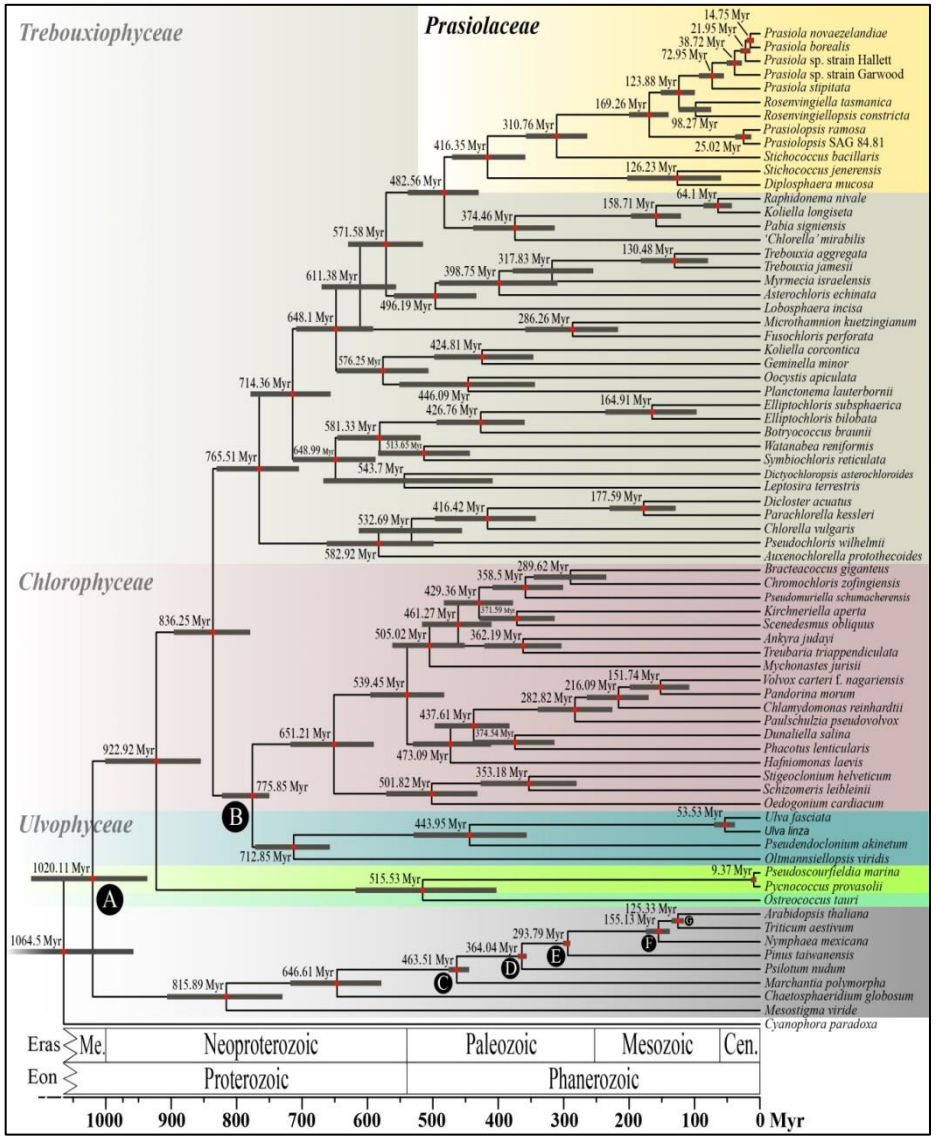
Tracer software

Fossil data



5 Ascomycete fossils and 4 markers

(Adapted from Beimforde et al. 2014)



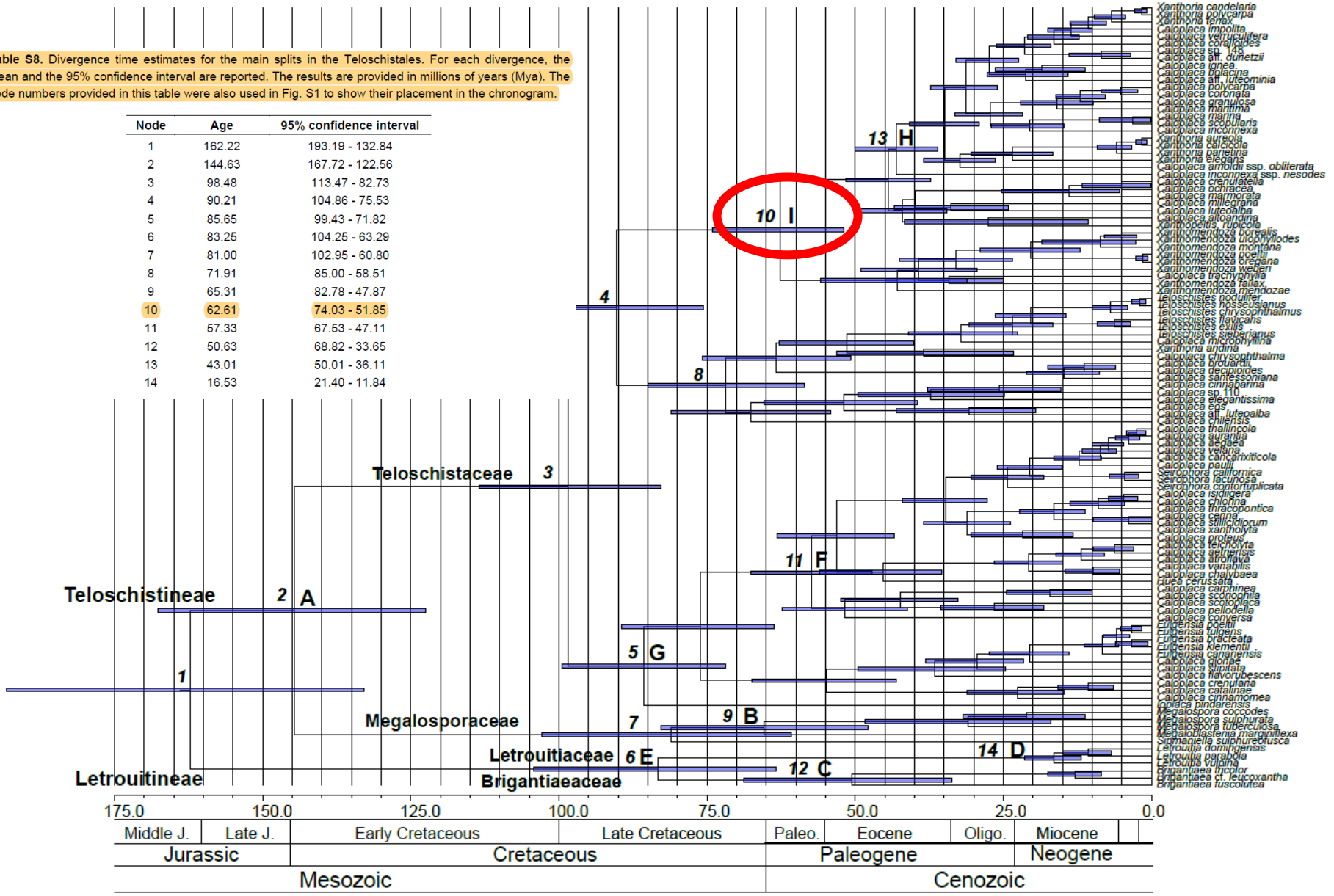
7 fossil evidences and 5 markers (7867 bp)

(Adapted from Garrido-Benavent et al. 2018)

Drawing an age estimate for a secondary calibration

Table S8. Divergence time estimates for the main splits in the Teloschistales. For each divergence, the mean and the 95% confidence interval are reported. The results are provided in millions of years (Mya). The node numbers provided in this table were also used in Fig. S1 to show their placement in the chronogram.

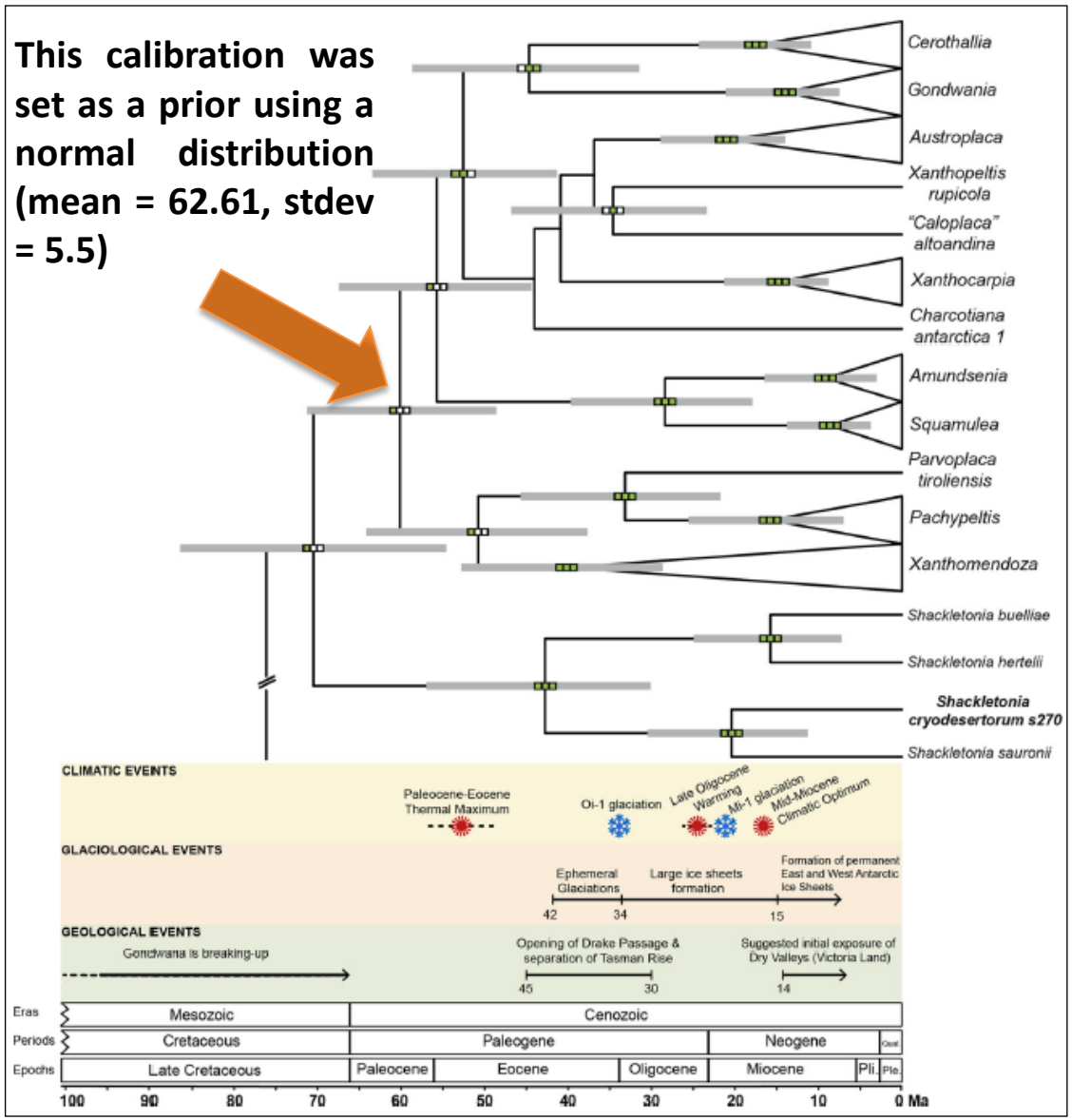
Node	Age	95% confidence interval
1	162.22	193.19 - 132.84
2	144.63	167.72 - 122.56
3	98.48	113.47 - 82.73
4	90.21	104.86 - 75.53
5	85.65	99.43 - 71.82
6	83.25	104.25 - 63.29
7	81.00	102.95 - 60.80
8	71.91	85.00 - 58.51
9	65.31	82.78 - 47.87
10	62.61	74.03 - 51.85
11	57.33	67.53 - 47.11
12	50.63	68.82 - 33.65
13	43.01	50.01 - 36.11
14	16.53	21.40 - 11.84



Adapted from Gaya et al. (2015)

Using that age estimate

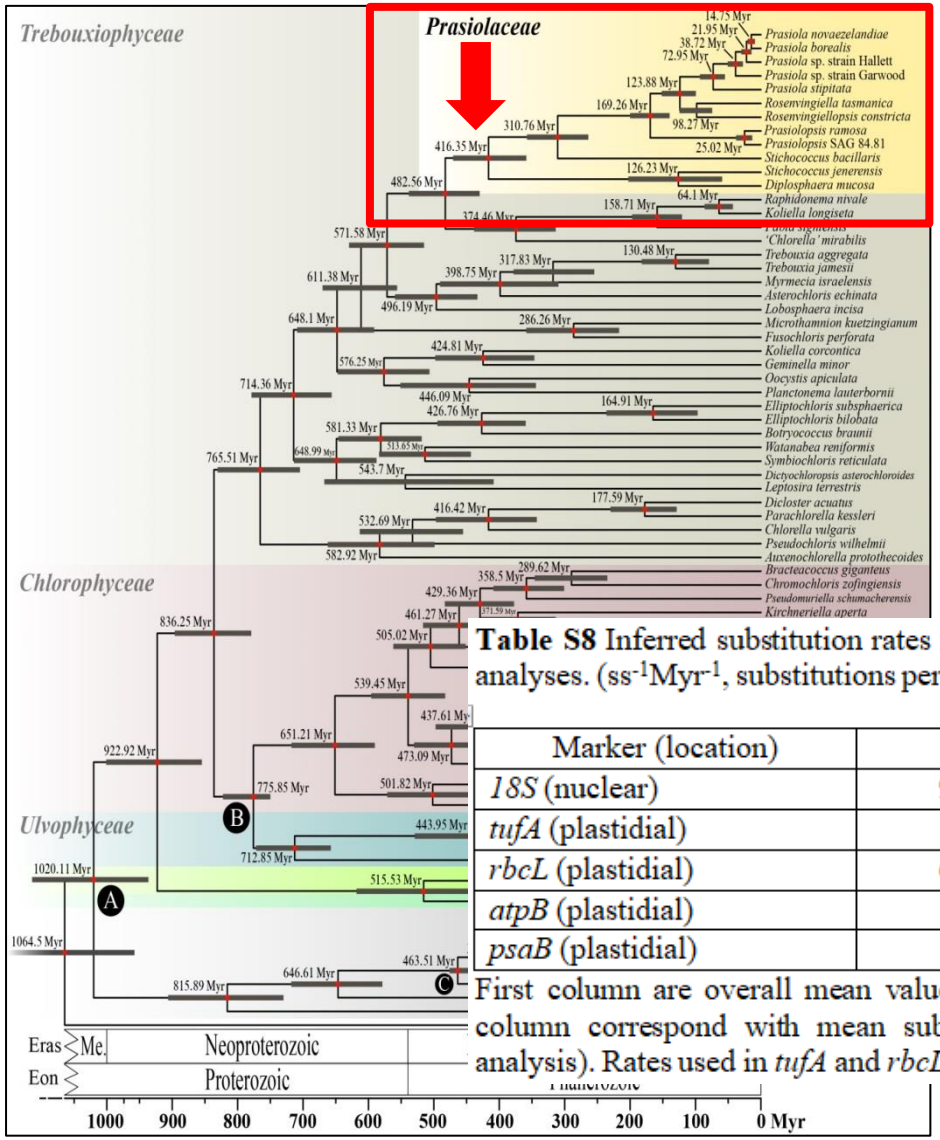
This calibration was set as a prior using a normal distribution (mean = 62.61, stdev = 5.5)



The dataset comprised three markers (ITS, nuLSU, mtSSU)

Adapted from Garrido-Benavent et al. (2016)

From the primary calibration we draw **substitution rates** for secondary calibrations



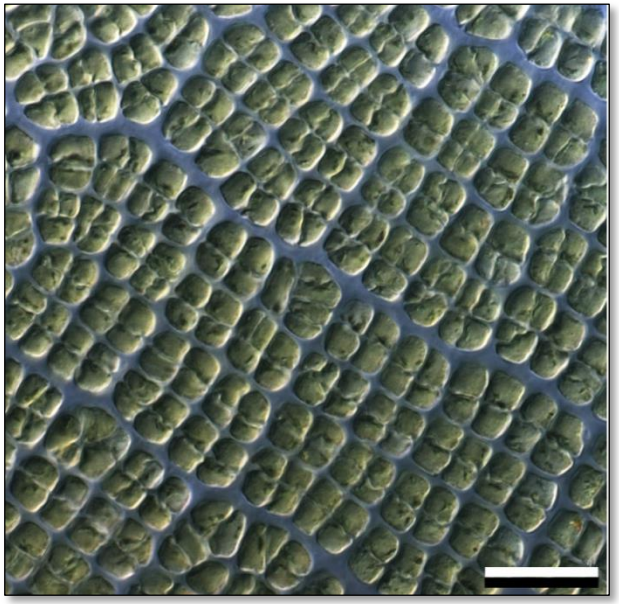


Table S8 Inferred substitution rates for selected markers from first- and second-step photobiont dating analyses. ($ss^{-1}Myr^{-1}$, substitutions per site per million years).

Marker (location)	Overall mean rate	*Prasiolaceae* mean rate
18S (nuclear)	9.416×10^{-5} $ss^{-1}Myr^{-1}$	5.05×10^{-5} $ss^{-1}Myr^{-1}$
tufA (plastidial)	1.109×10^{-3} $ss^{-1}Myr^{-1}$	**1.28×10^{-3} $ss^{-1}Myr^{-1}$**
rbcL (plastidial)	6.078×10^{-4} $ss^{-1}Myr^{-1}$	**9.57×10^{-4} $ss^{-1}Myr^{-1}$**
atpB (plastidial)	1.208×10^{-3} $ss^{-1}Myr^{-1}$	1.54×10^{-3} $ss^{-1}Myr^{-1}$
psaB (plastidial)	1.1×10^{-3} $ss^{-1}Myr^{-1}$	1.33×10^{-3} $ss^{-1}Myr^{-1}$

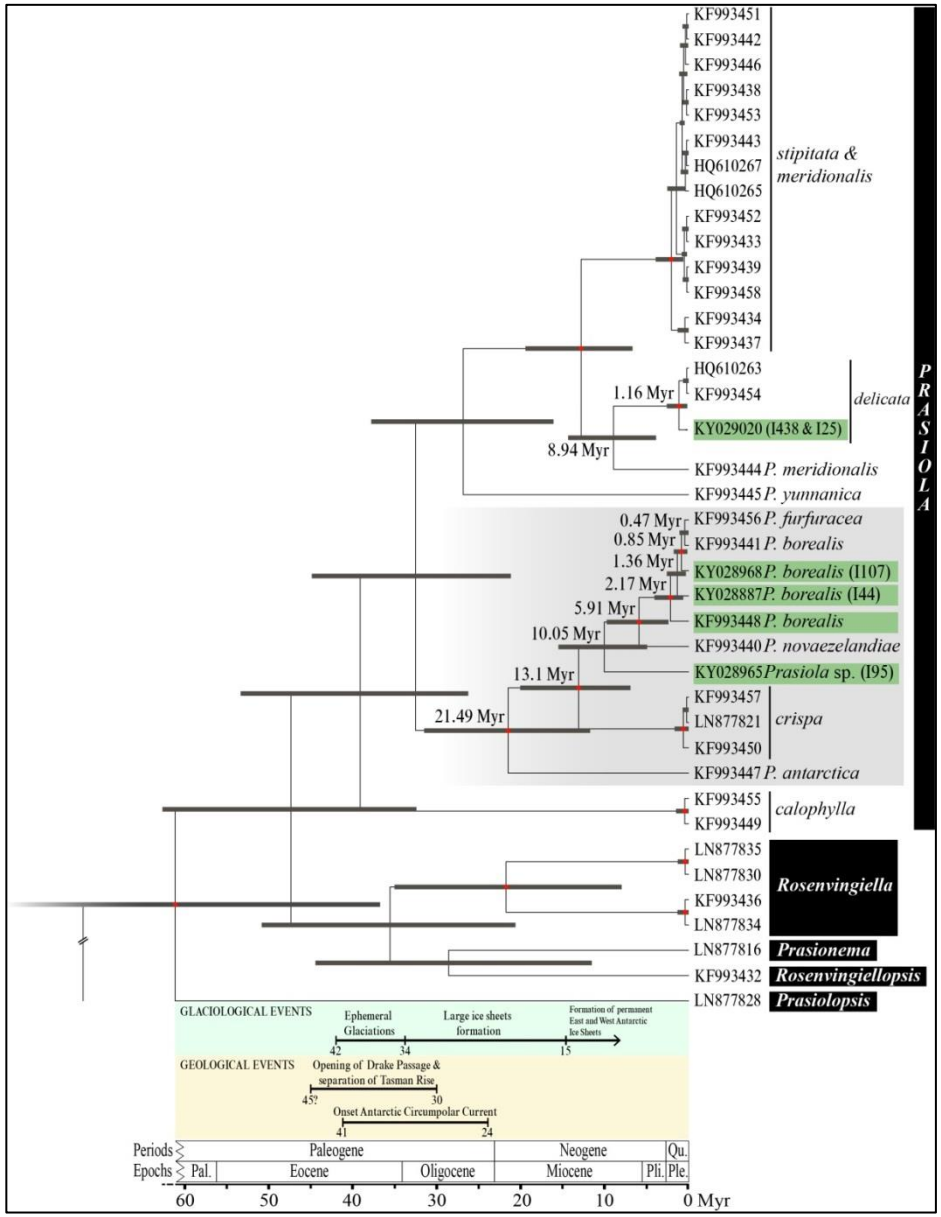
First column are overall mean values obtained in the first-step analysis, while values in the second column correspond with mean substitution rates drawn specifically from *Prasiolaceae* (first-step analysis). Rates used in *tufA* and *rbcL* single-locus dating analyses are in bold.

7 fossil evidences and 5 markers (7867 bp)

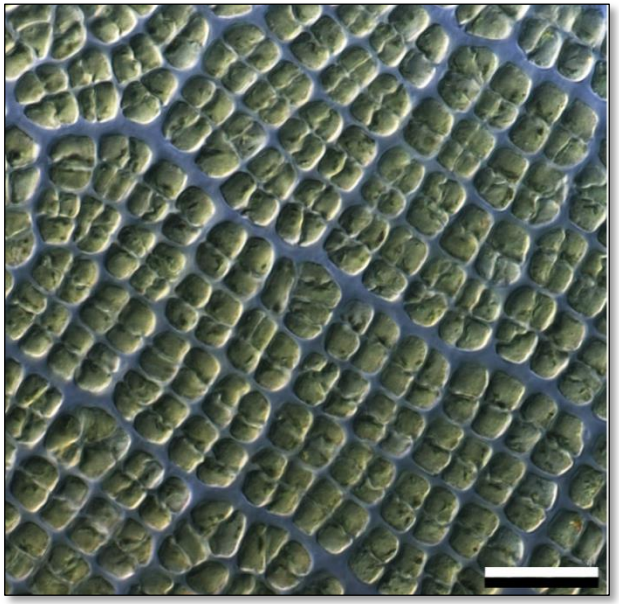
(Adapted from Garrido-Benavent et al. 2018)

FigTree

From the primary calibration we draw **substitution rates** for secondary calibrations



Adapted from Garrido-Benavent et al. (2018)

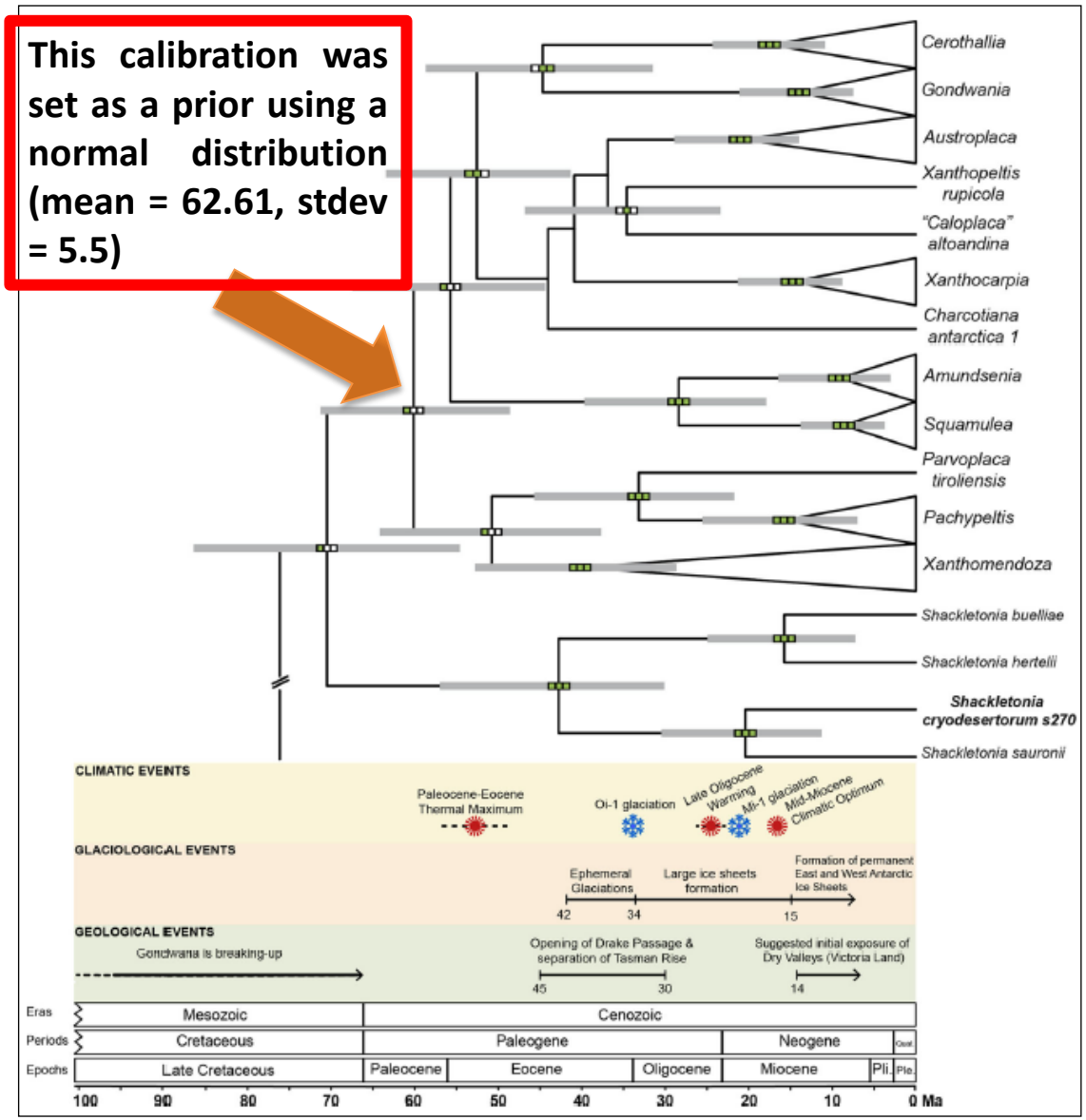


Chronogram inferred in BEAST from *tufA* data of selected ***Prasiolaceae*** members, including most ***Prasiola*** species

The estimated substitution rate for the *tufA* marker was used as a secondary calibration in an extended phylogeny of *Prasiolaceae*.

Comparing results of dating analyses

This calibration was set as a prior using a normal distribution (mean = 62.61, stdev = 5.5)



The dataset comprised three markers (ITS, nuLSU, mtSSU)

Substitution rates for ITS:

- 3.41×10^{-9} s/s/y (*Melanohalea*, Leavitt et al. 2012)
- 2.43×10^{-9} s/s/y (*Montanelia*, Leavitt et al. 2015)

Age estimates are known to be quite biased or inexact because of many reasons. In my opinion, the best approach is to apply two or three different calibrations to the same dataset, and report and discuss all the results at the same time. For example, I used here an age estimate for this clade as well as two different substitution rates for the ITS.

Adapted from Garrido-Benavent et al. (2016)

Comparing results of dating analyses

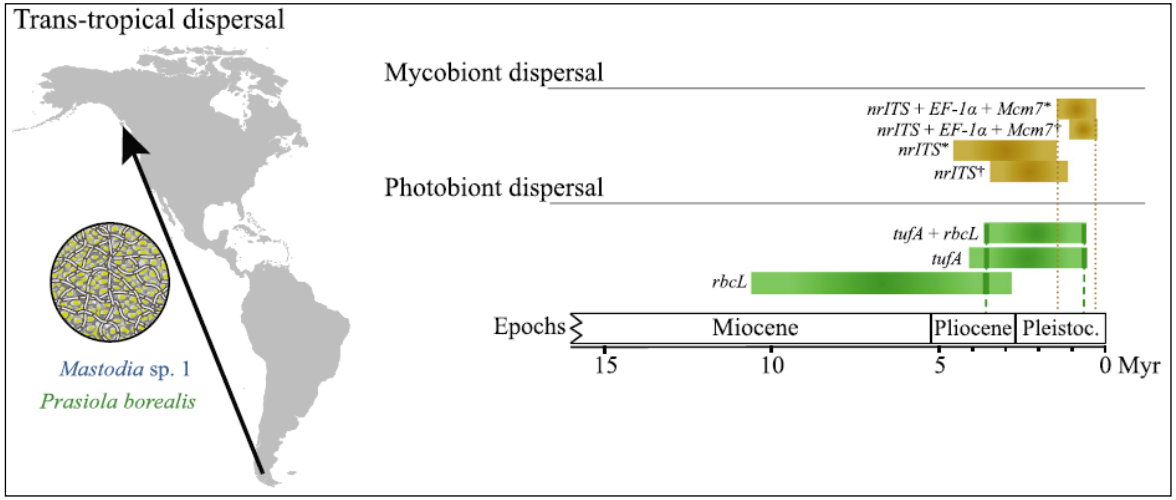
Table 4 Divergence time estimates (Ma) of selected nodes obtained using different secondary calibration approaches with BEAST

	Calibrated node approach	*Melanohalea* nrITS substitution rate[a]	*Montanelia* nrITS substitution rate[b]	Gaya et al. (2015)	Geological period
Shackletonia and remaining *Xanthorioideae* lineages split	70.5 (54.8–86.2)	64 (51.4–77.9)	89.3 (70.9–107.7)	–	Late Cretaceous-Paleogene
Shackletonia crown group	42.8 (30.3–56.7)	38.7 (28.1–50.1)	54.2 (39.4–71)	–	Late Paleocene-Early Oligocene
Shackletonia cryodesertorum origin	20.4 (11.5–30.2)	18.5 (10.7–27.4)	25.9 (14.7–37.6)	–	Late Oligocene-Early Miocene
Xanthocarpia crown group	14.79 (9–21.1)	13.5 (8.7–18.9)	18.8 (12.1–25.9)	*c.* 14 (5.5–25.5)	Early/Middle Miocene
Xanthomendoza crown group	40.12 (28.9–52.5)	36.4 (26.9–46.9)	50.9 (37–64.8)	*c.* 43 (31–56)	Eocene-Early Oligocene
Caloplaca-Xanthoria and *Xanthomendoza* split	60.15 (48.8–71)	54.2 (43–65.8)	76 (60–91.8)	62.61 (74.03–51.85)	Late Cretaceous-Paleogene

Proposed geological periods take into account estimated ages within 95 % HPD obtained from the first two analyses. Results of Gaya et al. (2015) are provided for comparison

[a] 3.41×10^{-9} s/s/y (Leavitt et al. 2012); [b] 2.43×10^{-9} s/s/y (Leavitt et al. 2015)

Adapted from Garrido-Benavent et al. (2016)



Adapted from Garrido-Benavent et al. (2018)

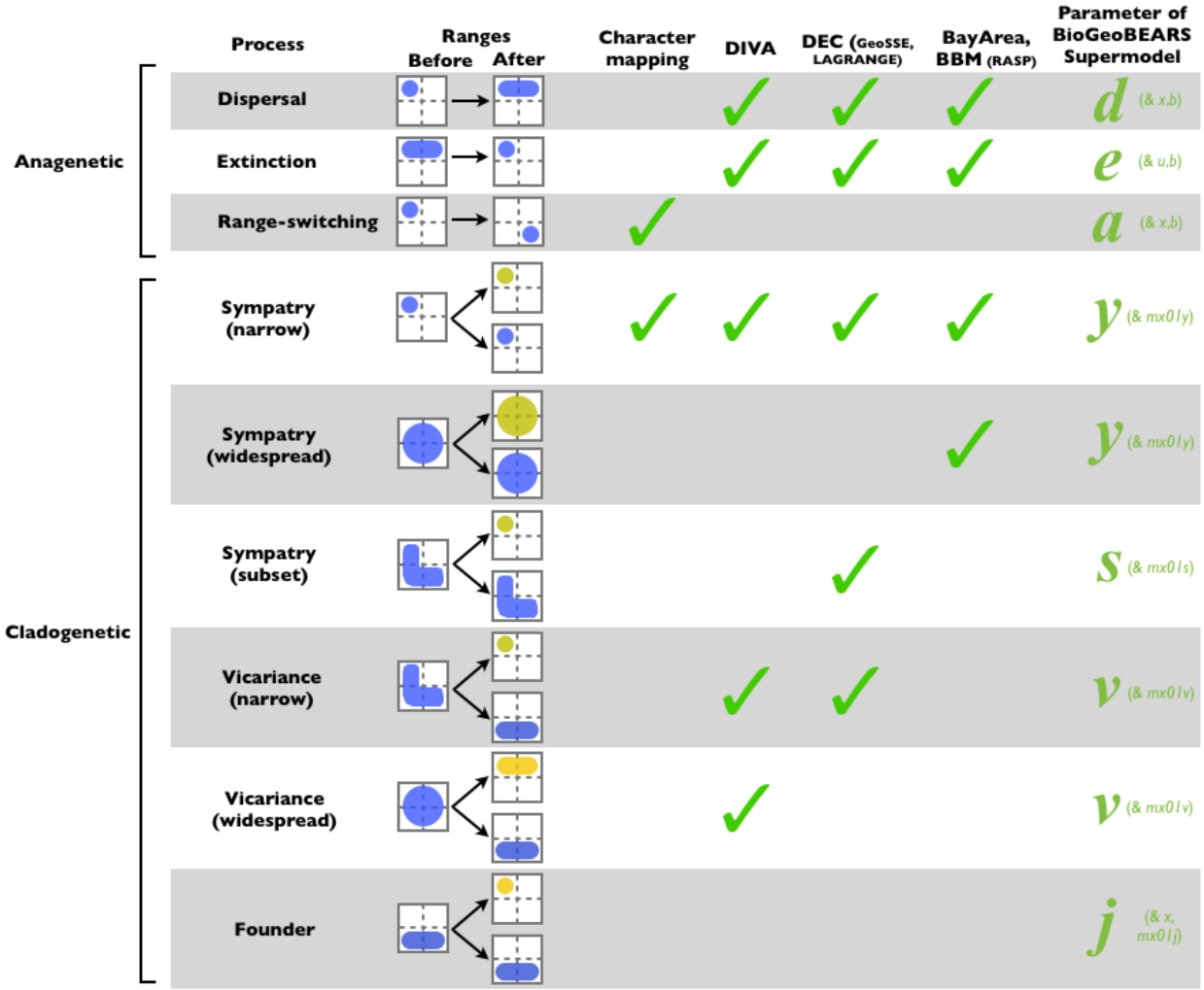
Substitution rates for the fungal *ITS*:

- 2.52×10^{-9} s/s/y (*Erysiphales*, Leavitt et al. 2012)
- 3.41×10^{-9} s/s/y (*Melanohalea*, Leavitt et al. 2015)

Substitution rates for the algal *tufA*:

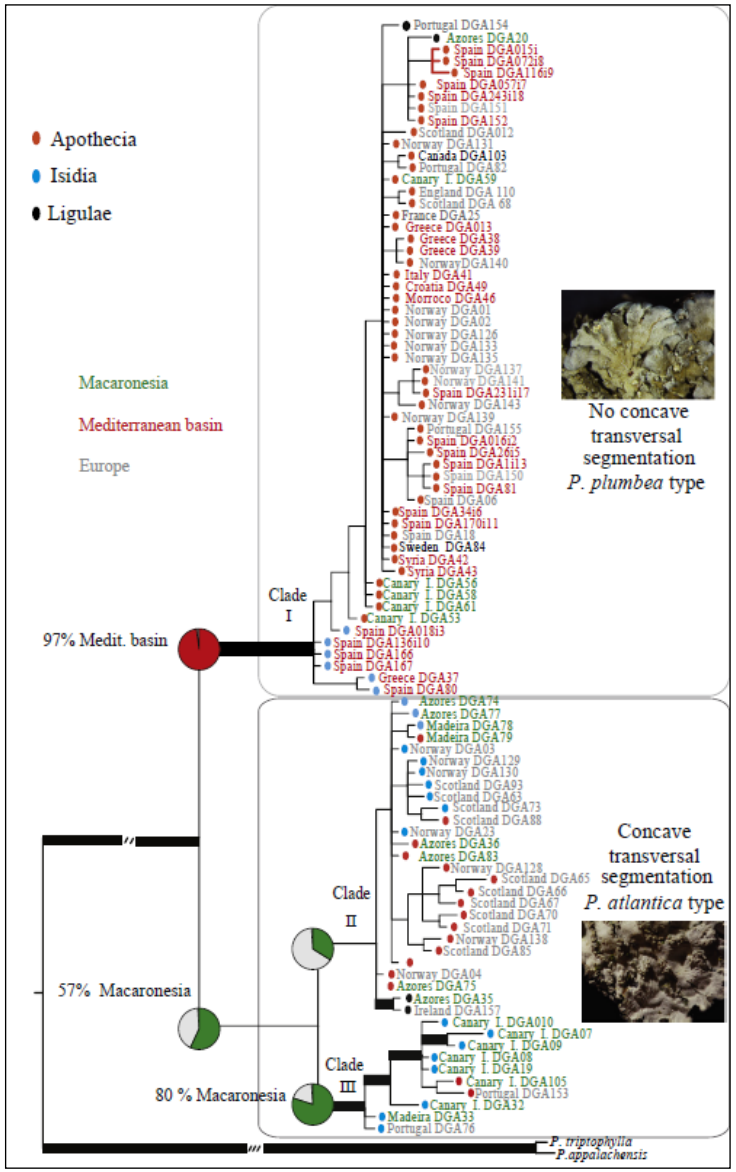
- 1.28×10^{-9} s/s/y (*Prasiolaceae*, Garrido-Benavent et al. 2018)

Ancestral range reconstruction



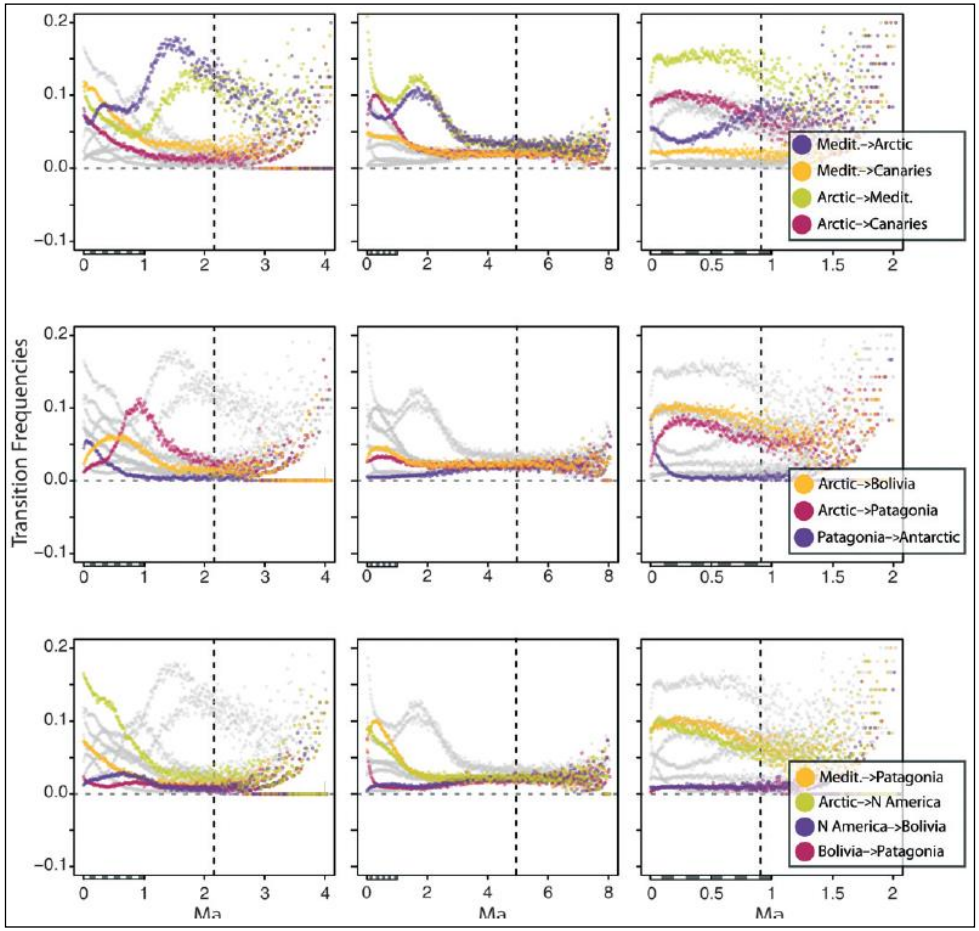
Reconstructing the ancestral range of species has become popular in many studies of lichens in the last years. There are several methods and programs that account for different anagenetic and cladogenetic processes that can alter the geographic range of a species such as dispersal, extinction or range switching, vicariance, founder event, and so on. Programs such as DIVA, Lagrange, BayArea, Biogeobears account for some of these processes.

Ancestral range reconstruction



Adapted from Otálora et al. (2017)

SIMMAP: stochastic character mapping of discrete traits on phylogenies (Huelsenbeck et al. 2003; Bollback 2006)

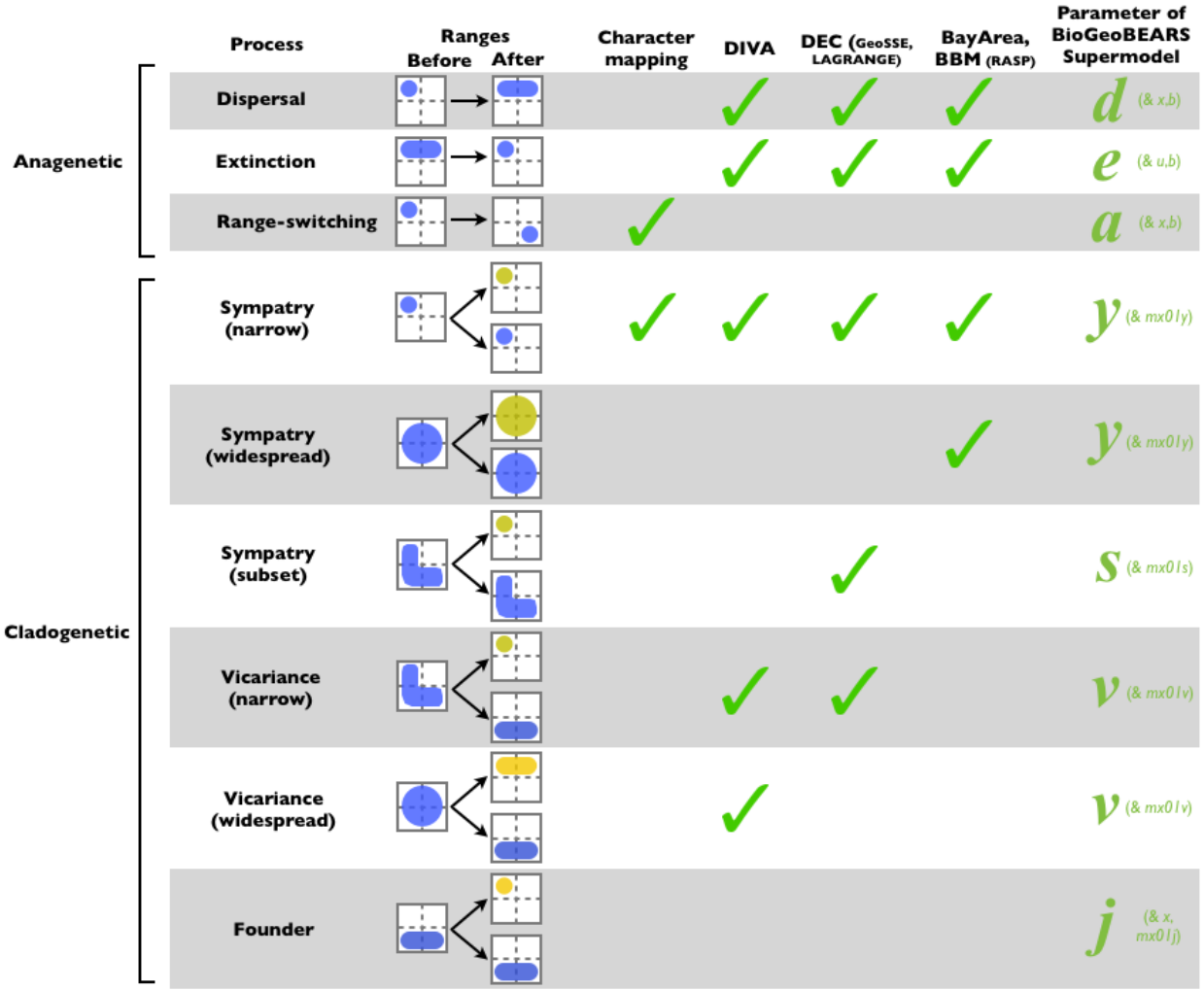


Adapted from Fernández-Mendoza & Printzen (2013)

R package *phytools*, function *make.simmap*

Ancestral range reconstruction

BioGeoBEARS (Matzke 2013)



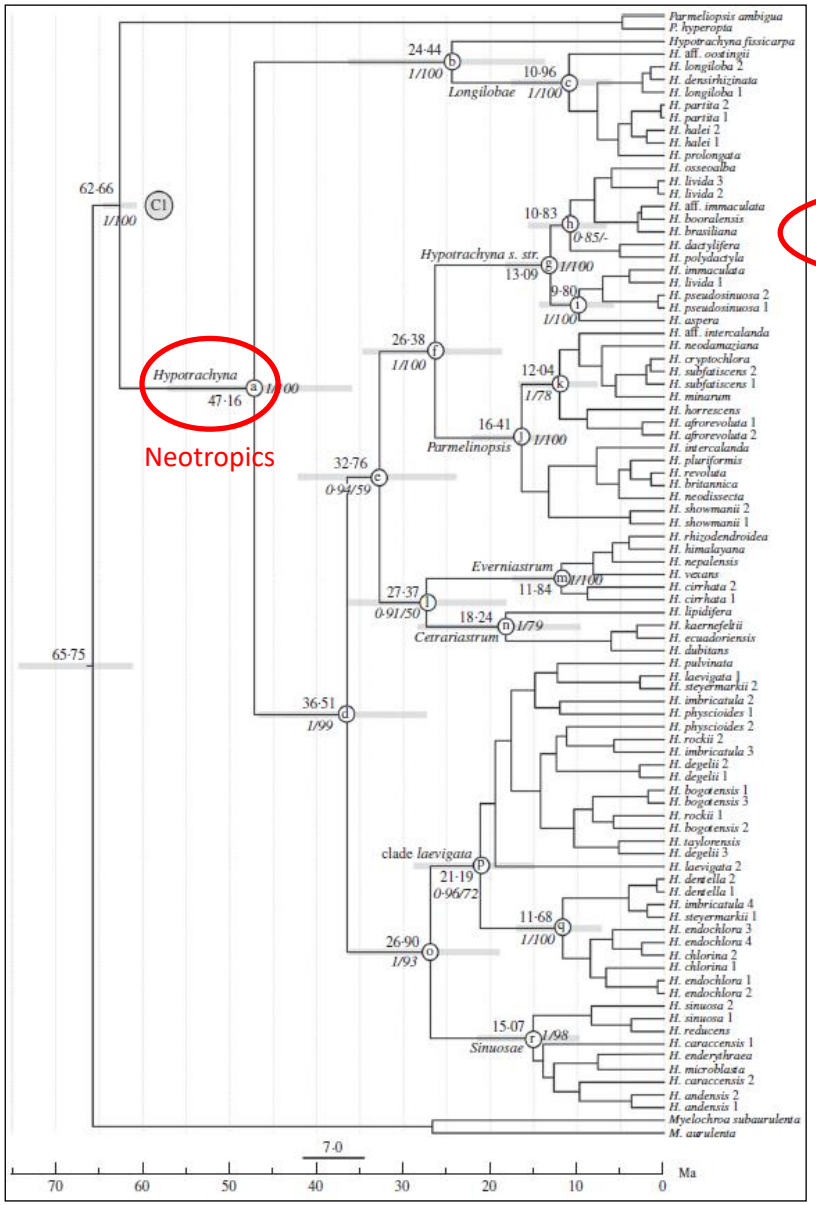
DEC
DEC+J

DIVALIKE
DIVALIKE+J

BAYAREALIKE
BAYAREALIKE+J

package *BioGeoBEARS*

Ancestral range reconstruction



Historical biogeography of ***Hypotrachyna*** (*Parmeliaceae*)

Node	Clade	Ancestral range probabilities								
		D	E	G	CD	DE	DG	EG	CDF	DEG
a	*Hypotrachyna*	**0·20**	0·02	0·03	0·00	0·05	0·05	0·01	0·00	0·01
b	subgen. *Longilobae*/*H. fissicarpa*	0·04	**0·12**	0·04	0·00	**0·22**	0·01	**0·29**	0·00	**0·16**
c	subgen. *Longilobae*	0·01	0·00	**0·19**	0·00	0·00	**0·58**	0·05	0·00	**0·13**
d	all species except subgen. *Longilobae*/*H. fissicarpa*	**0·28**	0·00	0·00	0·01	0·01	0·02	0·00	0·00	0·00
e		**0·72**	0·00	0·00	0·02	0·01	0·01	0·00	0·00	0·00
f	subgen. *Hypotrachyna* + *Parmelinopsis*	**0·47**	0·00	0·00	0·02	0·02	0·01	0·00	0·00	0·00
g	*Hypotrachyna* s. str.	**0·77**	0·00	0·00	0·02	0·01	0·01	0·00	0·00	0·00
h		**0·77**	0·00	0·00	0·02	0·02	0·02	0·00	0·00	0·00
i		**0·86**	0·00	0·00	0·01	0·01	0·01	0·00	0·00	0·00
j	subgen. *Parmelinopsis*	0·02	0·00	0·00	0·00	0·01	0·00	0·00	0·00	0·00
k		0·01	0·02	0·00	0·00	0·00	0·00	0·00	0·00	0·00
l	subgen. *Cetrariastrum* + *Everniastrum*	**0·84**	0·00	0·00	0·07	0·01	0·00	0·00	0·00	0·00
m	subgen. *Everniastrum*	**0·11**	0·00	0·00	**0·26**	0·01	0·00	0·00	**0·14**	0·00
n	subgen. *Cetrariastrum*	**0·99**	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
o	*H. laevigata* group + subgen. *Sinuosae*	**0·28**	0·00	0·00	0·01	0·01	0·01	0·00	0·00	0·00
p	*H. laevigata* group	**0·85**	0·00	0·00	0·01	0·00	0·01	0·00	0·00	0·00
q		**0·58**	0·00	0·00	0·03	0·01	0·04	0·00	0·01	0·00
r	subgen. *Sinuosae*	0·09	0·00	0·00	0·01	0·01	0·01	0·00	0·00	0·00

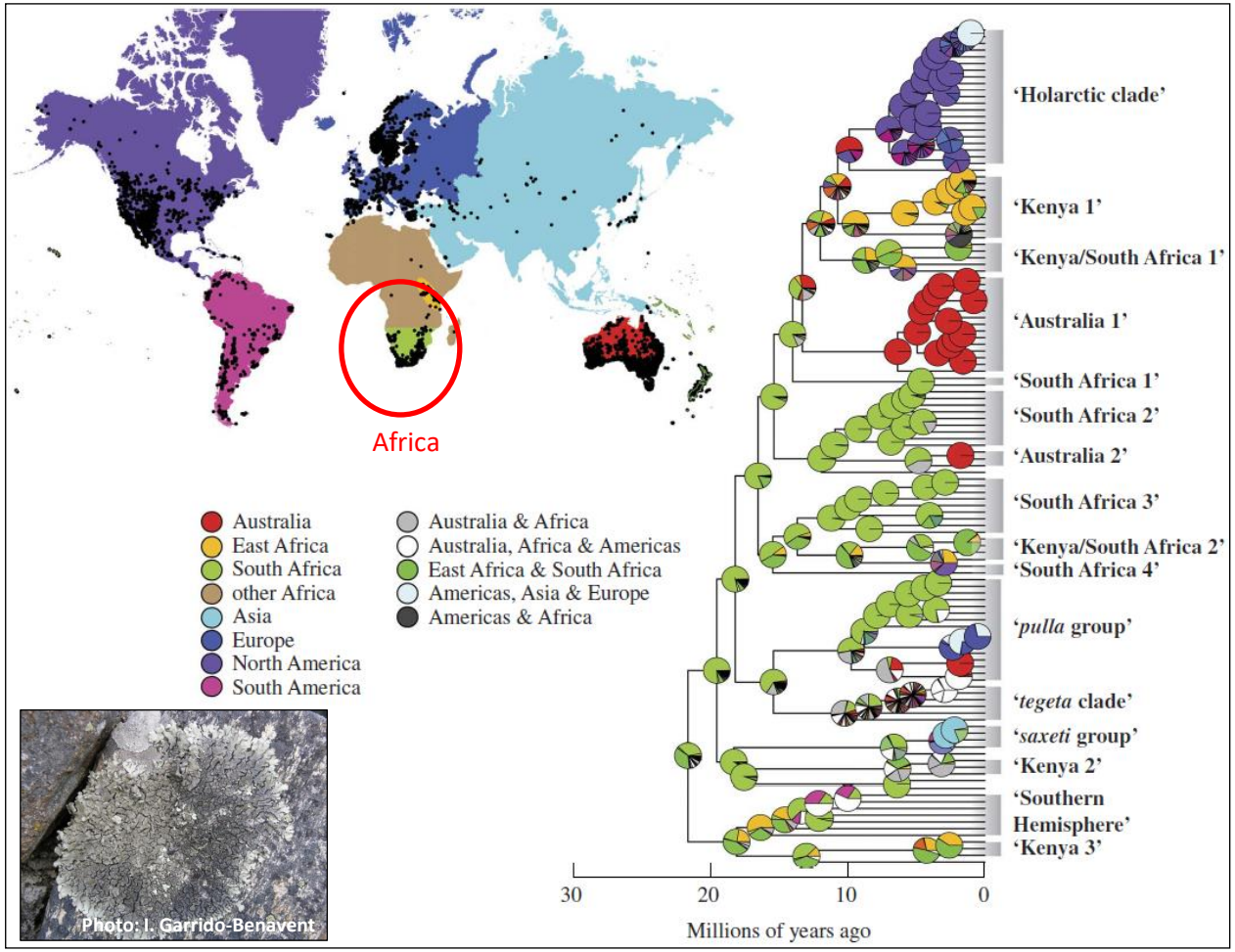
Adapted from Cubas et al. (2018)

Ancestral range reconstruction

Historical biogeography of ***Xanthoparmelia*** (*Parmeliaceae*)

Model	LnL	Number of parameters	Parameter estimates *d*	*e*	*j*	AIC
DEC	−391·1326	2	0·009079	3·313e-03	0·0	786·2652
DEC+J	−386·7134	3	0·007892	1·0 e-12	0·011631	779·4268

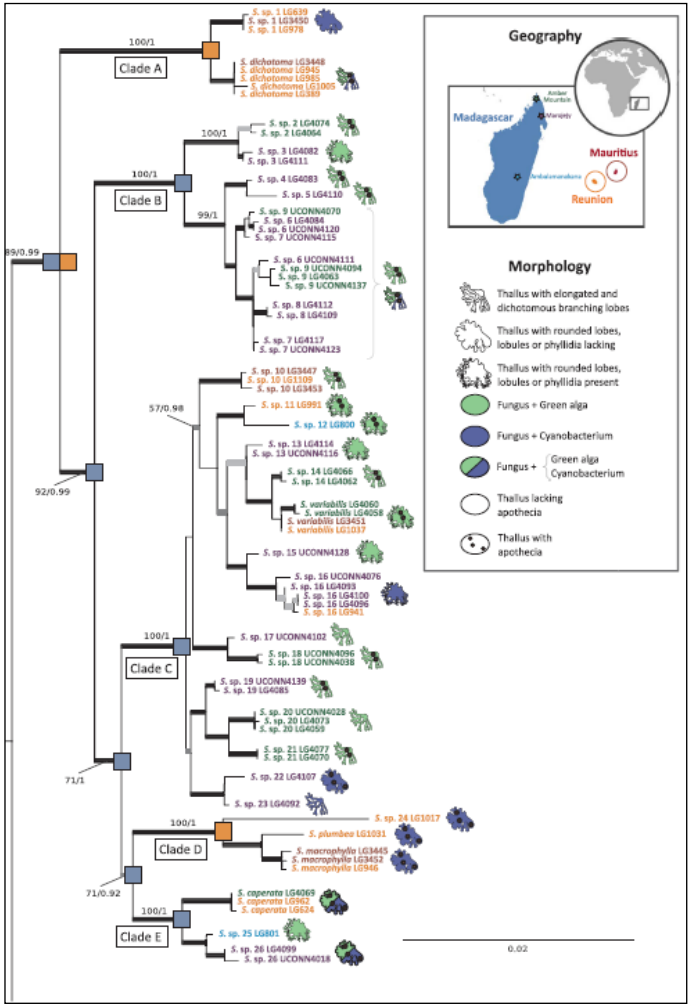
Parameters: the rate of range expansion ("dispersal"), parameter *d*; range contraction ("extinction"), parameter *e*; weight of each jump dispersal event in the cladogenesis matrix, parameter *j*. Akaike Information Criterion, AIC.



Adapted from Leavitt et al. (2018)

Ancestral range reconstruction

	Model	Ln L	AICc	P-value (LRT)	Most likely ancestral range
Three geographic regions (Madagascar, Mauritius, Réunion)	DEC	−66.46	137.1	$2.5\ 10^{-9}$	Madagascar + Réunion
	DEC + J	−48.69	103.7		**Madagascar + Réunion**
	DIVA	−66.04	136.3	$2.6\ 10^{-9}$	Madagascar + Réunion
	DIVA + J	−48.31	103		**Madagascar + Réunion**
	BAYAREALIKE	−114.2	232.7	$7.8\ 10^{-30}$	Madagascar + Réunion
	BAYAREALIKE + J	−49.88	106.1		**Réunion**



Historical biogeography of ***Sticta*** (*Lobariaceae*) in Madagascar and the Mascarenes

Adapted from Simon et al. (2018)

Ancestral range reconstruction

DOI: 10.1111/jbi.13173

PERSPECTIVE

WILEY Journal of Biogeography

Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection

Ree & Sanmartín (2018)

	Model	Ln L	AICc	P-value (LRT)	Most likely ancestral range
Three geographic regions (Madagascar, Mauritius, Réunion)	DEC	−66.46	137.1	2.5 10^{-9}	Madagascar + Réunion
	DEC + J	−48.[illegible]	103.7		Madagascar + Réunion
	DIVA	−6[illegible].4	136.3	2.6 10^{-9}	Madagascar + Réunion
	DIVA + J	−48.[illegible]	103		Madagascar + Réunion
	BAYAREALIKE	−114.2	232.7	7.8 10^{-30}	Madagascar + Réunion
	BAYAREALIKE + J	−49.88	106.1		Réunion

Likelihoods from DEC and DEC+J are not statistically comparable

"…For simple inference of ancestral ranges on a fixed phylogeny, a DEC-based model may be defensible if statistical model selection is not used to justify the choice…"

"…If different models confidently yield conflicting reconstructions that are meaningful to the study, it seems entirely reasonable to favour one over another by making arguments based on empirical (biological, geographic) considerations—in other words, to allow non-statistical judgements guide model choice…"

SOLUTION: select any model [DEC(+J), DIVALIKE(+J), BAYAREALIKE(+J)] or more than one model, and discuss results under the prism of our study group (discuss which is more biologically realistic). BUT NEVER COMPARE THE LIKELIHOODS OF THESE MODELS!!!!!

What if…

…we are dealing with only

1 species

BioGeoBEARS: treat each distinct lineage as a separate species

SIMMAP

BEAST: discrete phylogeography analysis (Lemey et al.)

…we are dealing with only

≥ 2 species

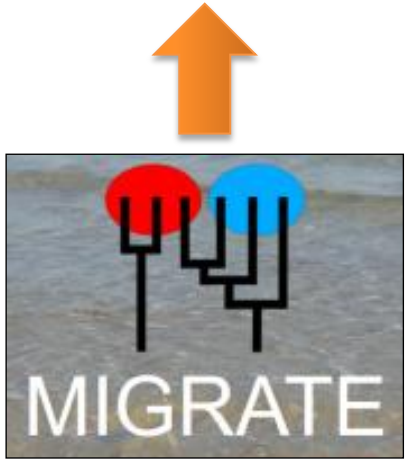
BioGeoBEARS: if there is strong population substructure within nominal species, or there are non-monophyletic species, treat each lineage as a separate putative species

SIMMAP

For more details see BioGeoBEARS, SIMMAP and BEAST discrete phylogeography tutorials

Migration

To obtain mutation-scaled immigration rate (***M***) and effective population size (***Θ***) estimates

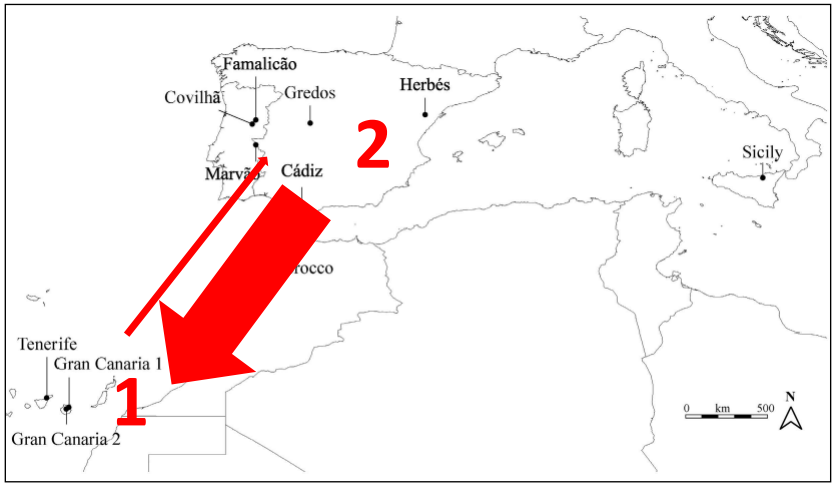


Beerli (2006)
Beerli & Palczewski (2010)

The last interesting biogeographic analysis is migration, which can be conducted with a coalescent-based method in MIGRATE. First, this software can be used to estimate mutation-scaled immigration rates and effective population sizes.

Migration

To assess the direction and intensity of gene flow



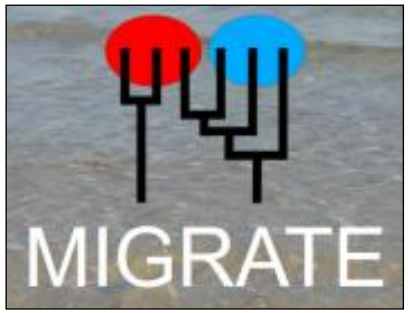
Parameter	2.5%	25.0%	75.0%	97.5%	Mean
Θ1	0.17	0.73	1.53	2.1	1.16
Θ2	9.13	11.7	15.9	21.33	16.59
M2->1	3.2	3.97	5.07	5.7	8.39
M1->2	0.03	0.53	1.3	1.83	0.94

Adapted from Alors et al. (2017)

For example, in a phylogeographic study of Alors et al. about *Parmelina carporrhizans*, they inferred higher effective population sizes in the Mediterranean populations than in the Macaronesian. The intensity of gene flow was also higher from the Mediterranean towards the Canary Islands than viceversa.

Migration

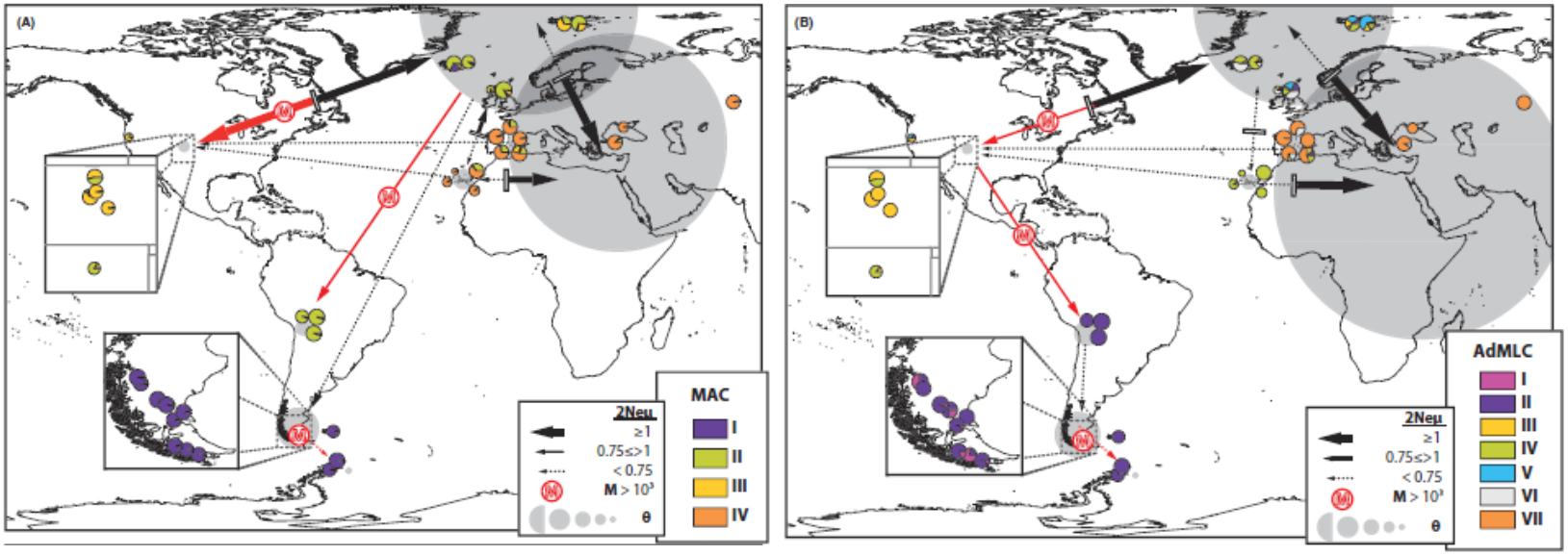
To obtain mutation-scaled immigration rate (***M***) and effective population size (***Θ***) estimates



Beerli (2006)
Beerli & Palczewski (2010)

To simulate and compare different models of population structure and gene flow

Migration



Model name	Structure and number of parameters (θ + *M*)		Description	Bezier Lml	Model probability	2 ln BF Bezier
NULL models						
NULL	No connections	1 + 0	Single admixed population	−5784.19	1.89×10^{-937}	4312.48
FULL	All connections	13 + 156	Full 13 populations model	−4214.73	7.67×10^{-256}	1173.56
Burst dispersal from the Arctic						
1.1	Arc→[Med,Can,NAm, Bol,Pat]→Ant	7 + 6	Full N-S directionality	−3685.91	3.54×10^{-26}	115.92
1.2	Arc→[Med,Can,NAm, Bol,Pat]; Pat→Ant	7 + 6	Radiation from the arctic	−3693.91	1.19×10^{-29}	131.92
1.3	Arc↔Med, NAm→Arc; Arc→[Can,Bol,Pat]; Pat→Ant	7 + 7	Temperate refuge, Burst from the Arctic, Exchange with the Mediterranean	−3688.27	3.34×10^{-27}	120.64
A (1.4)	((Med↔Can) ↔Arc); [Med, Can]–> NAm;Arc→Pat→Ant; Arc→Bol; Arc→NAm	7 + 13	Radiation from the Arctic, High Northern Hemisphere connectivity.	−3627.95	0.525	0.00

Adapted from Fernández-Mendoza & Printzen (2013)

Migration

Biont	Directionality	Model	Description and number of estimated parameters (Θ + M)	Bézier lml	Model Probability	2ln BF Bézier
Mycobionts (*Mastodia* sp. 1 & sp. 2)	n/a	Null	No connections (single population) (1 + 0)	−5076.57	~0	2541.08
		Null	All possible connections among the 4 populations (4 + 12)	−4656.34	~0	1700.62
Bipolar mycobiont (*Mastodia* sp. 1) 	S→N	5	TF migration into NA. ANT migration into TF. (4 + 2)	−3829.29	7.897×10^{-11}	46.52
		6	**TF migration into NA and ANT. (4 + 2)**	**−3806.03**	**0.998**	**0**
		7	TF migration into NA and ANT. ANT migration into TF. (4 + 3)	−3817.84	7.416×10^{-6}	23.62
	N→S	8	NA migration into TF. TF migration into ANT. (4 + 2)	−3821.82	1.386×10^{-7}	31.58
		9	NA migration into TF. ANT migration into TF. (4 + 2)	−3960.56	7.72×10^{-68}	309.06
		10	NA migration into TF. TF migration into ANT and *vice versa* (4 + 3)	−3832.12	4.66×10^{-12}	52.18
Photobionts (*Prasiola borealis* & *Prasiola* sp.)	n/a	Null	No connections (single population) (1 + 0)	−4424.57	~0	845.62
		Null	All possible connections among the 4 populations (4 + 12)	−4287.1	~0	570.68
Bipolar photobiont (*Prasiola borealis*) 	S→N	5	TF migration into NA. ANT migration into TF. (4 + 2)	−4021.24	3.467×10^{-9}	38.96
		6	TF migration into NA and ANT. (4 + 2)	−4014.01	4.785×10^{-6}	24.5
		7	**TF migration into NA and ANT. ANT migration into TF. (4 + 3)**	**−4001.76**	**9.999×10^{-1}**	**0**
	N→S	8	NA migration into TF. TF migration into ANT. (4 + 2)	−4012.32	2.593×10^{-5}	21.12
		9	NA migration into TF. ANT migration into TF. (4 + 2)	−4103.53	6.336×10^{-45}	203.54
		10	NA migration into TF. TF migration into ANT and *vice versa* (4 + 3)	−4017.45	1.534×10^{-7}	31.38

Adapted from Garrido-Benavent et al. (2018)

Migration

NULL MODEL (all connections/populations=regions) — Full 4 populations model

NULL_full	Namerica	Chile	AntChil	Antart
NAmerica	Θnam	*	*	*
Chile	*	Θc	*	*
AntChil	*	*	Θanch	*
Antartida	*	*	*	Θan

nº calculated parameters: Θ= 4; M= 12

(**********)

1.1. Chile together with AntChil gives to Namer. Antarctica isolated.

MOD1.1	Namerica	Chile-AntChil	Antart
Namerica	Θnam	*	0
Chile-AntChil	0	Θchant	c
Antartida	0	c	Θan

nº calculated parameters: Θ= 3; M= 1

(**00*c0c*)

1.2. Chile together with AntChil gives to Namer. Antarctica gives to Chile.

MOD1.2	Namerica	Chile-AntChil	Antart
Namerica	Θnam	*	0
Chile-AntChil	0	Θchant	*
Antartida	0	0	Θan

nº calculated parameters: Θ= 3; M= 2

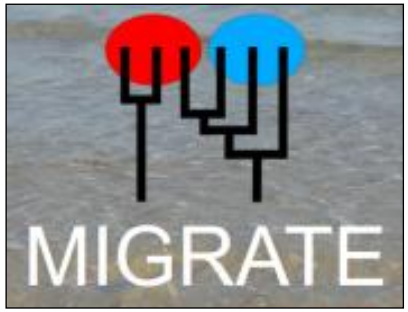
(**00**00*)

*: parameter that varies freely

c: a parameter with a fixed value

Migration

To obtain mutation-scaled immigration rate (***M***) and effective population size (***Θ***) estimates

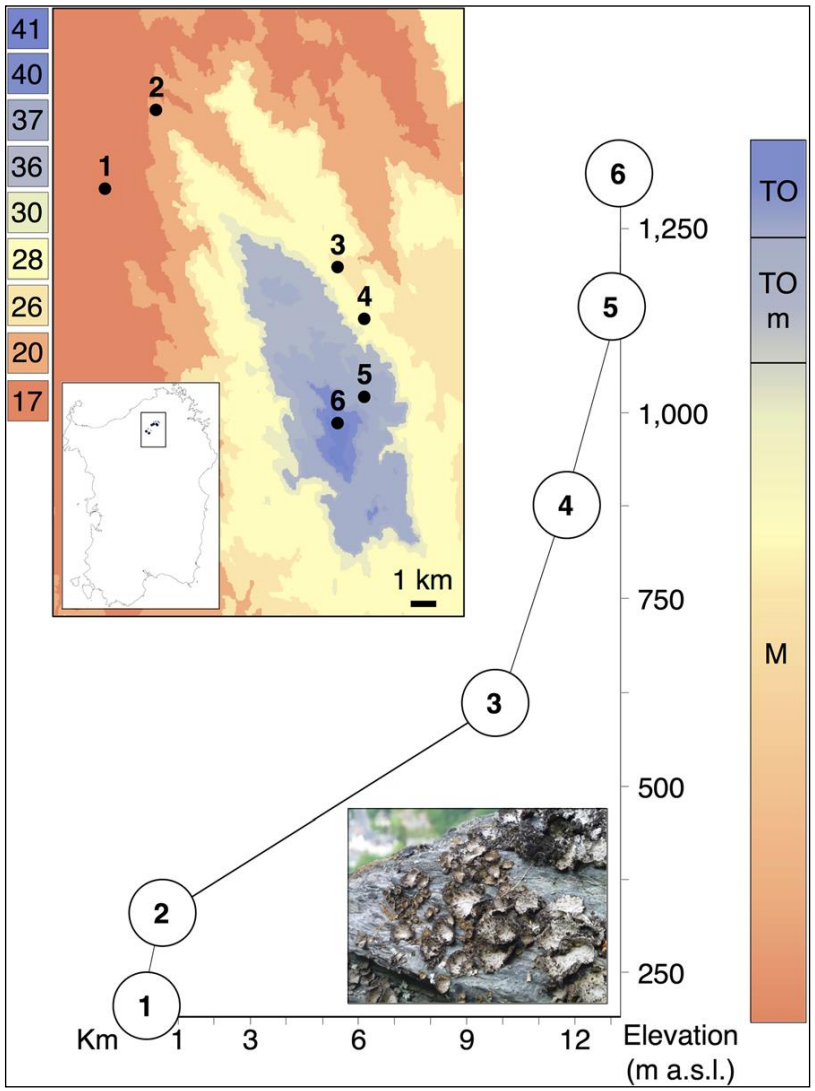


Beerli (2006)
Beerli & Palczewski (2010)

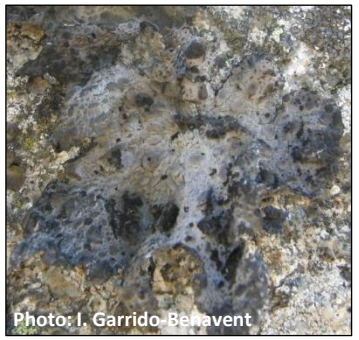
To simulate and compare different hypotheses of population structure and gene flow

To corroborate hypotheses of admixture

Migration



Adapted from Dal Grande et al. (2017)



Pop. 6

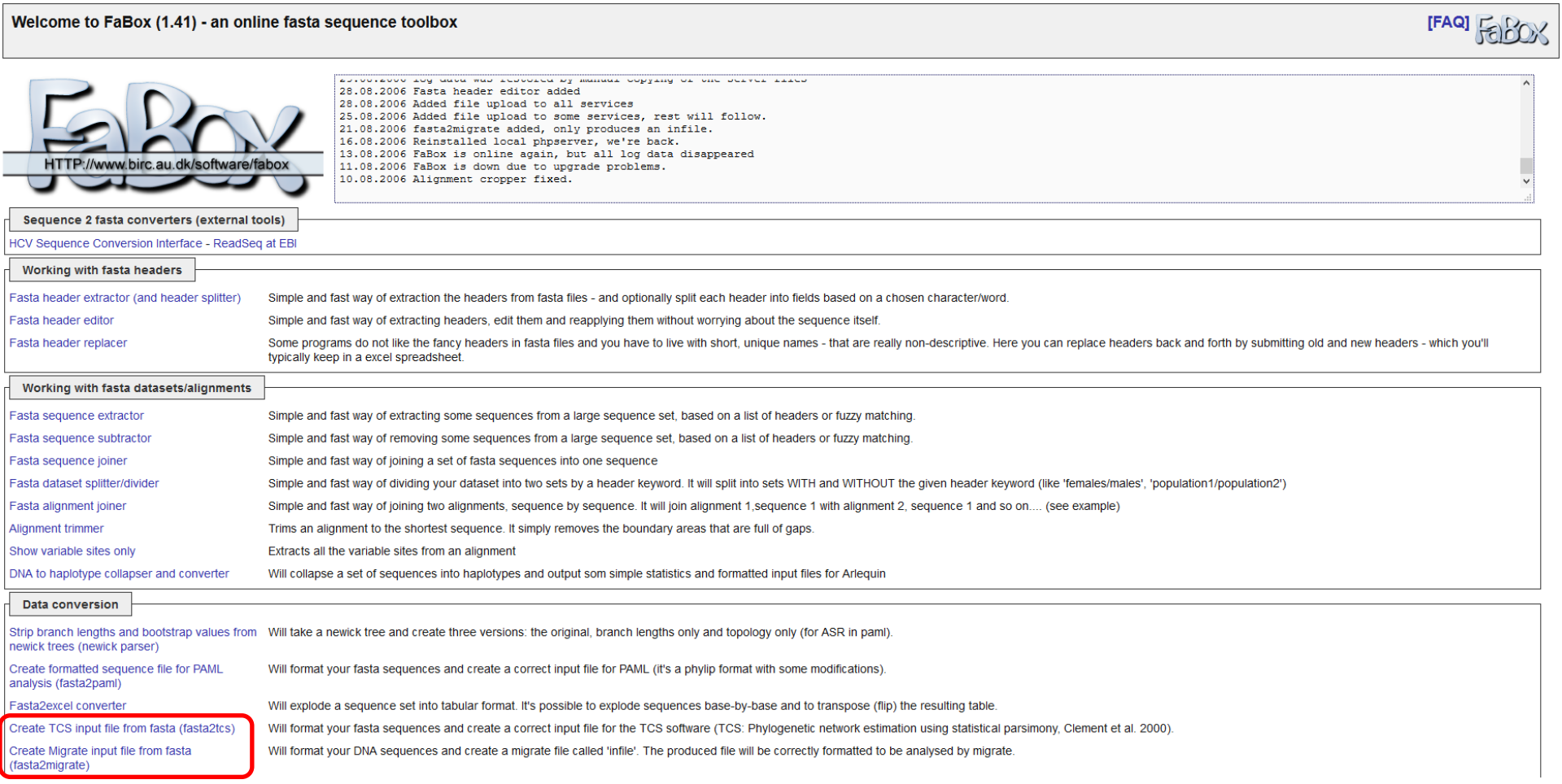
X

Pop. 5

admixed

Populations 1–4

FaBox is useful for building input files



http://users-birc.au.dk/biopv/php/fabox/

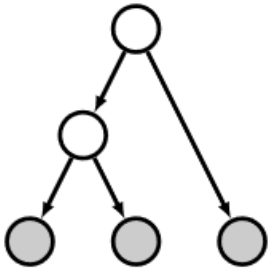
Villesen (2007)

Many analyses can be run online in CIPRES

RevBayes

Höhna, Sebastian, et al. "RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language." *Systematic Biology* 65.4 (2016): 726-736.

Software Tutorials Workshops Jobs Developer



RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

About

RevBayes provides an interactive environment for statistical computation in phylogenetics. It is primarily intended for modeling, simulation, and Bayesian inference in evolutionary biology, particularly phylogenetics. However, the environment is quite general and can be useful for many complex modeling tasks.

RevBayes uses its own language, Rev, which is a probabilistic programming language like JAGS, STAN, Edward, PyMC3, and related software. However, phylogenetic models require inference machinery and distributions that are unavailable in these other tools.

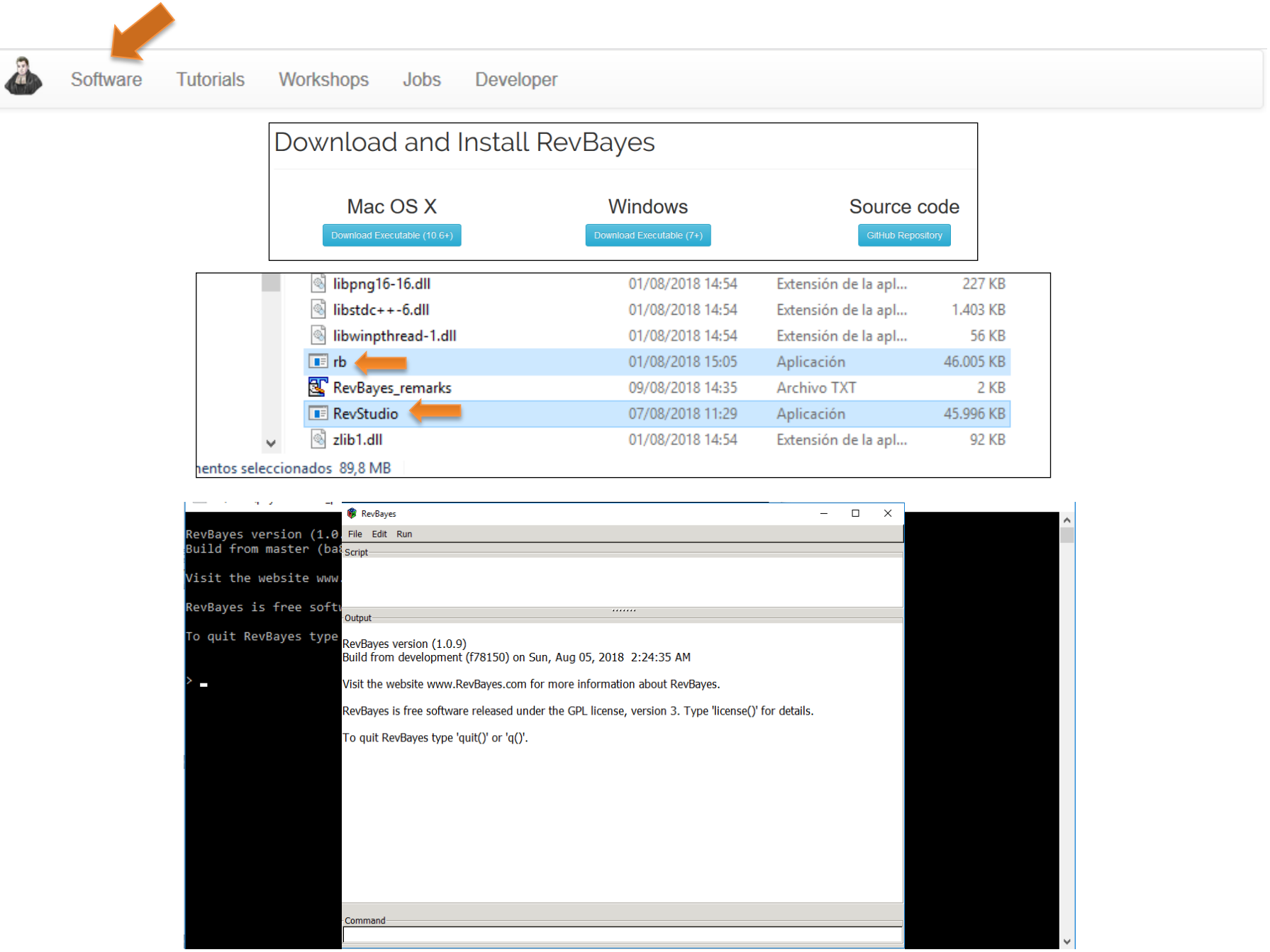
The Rev language is similar to the language used in R. Like the R language, Rev is designed to support interactive analysis. It supports both functional and procedural programming models, and makes a clear distinction between the two. Rev is also more strongly typed than R.

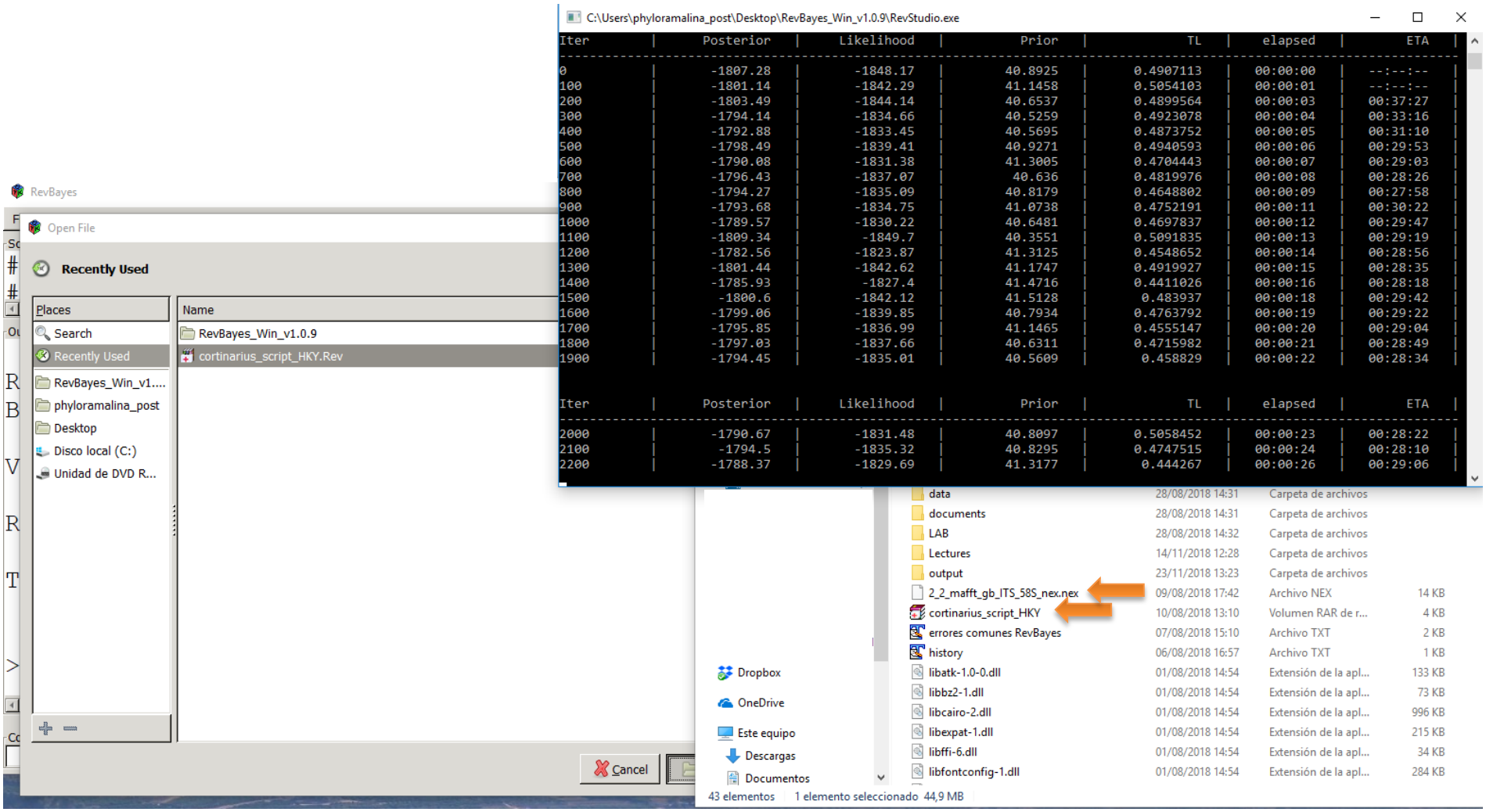
Core Development Team

RevBayes was designed and developed by Sebastian Höhna, Fredrik Ronquist and John P. Huelsenbeck. The core development team additionally includes Michael J. Landis, Bastien Boussau, Tracy A. Heath, Nicolas Lartillot, Walker Pett, and William A. Freyman.

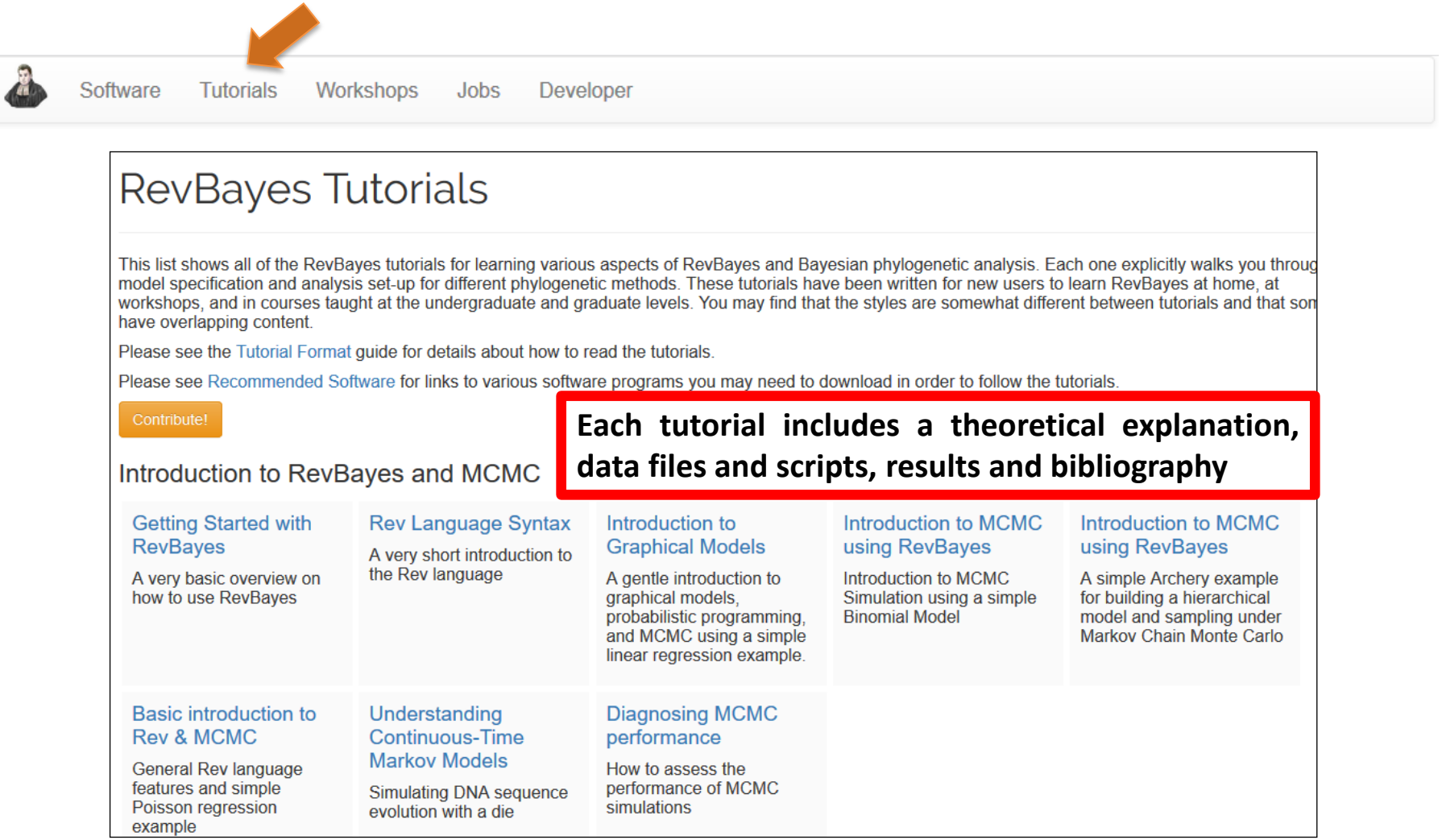
GitHub | License | Citation | Users Forum

https://revbayes.github.io/


Software
Tutorials
Workshops
Jobs
Developer
Download and Install RevBayes
Mac OS X
Download Executable (10.6+)
Windows
Download Executable (7+)
Source code
GitHub Repository
libpng16-16.dll 01/08/2018 14:54 Extensión de la apl... 227 KB
libstdc++-6.dll 01/08/2018 14:54 Extensión de la apl... 1.403 KB
libwinpthread-1.dll 01/08/2018 14:54 Extensión de la apl... 56 KB
rb 01/08/2018 15:05 Aplicación 46.005 KB
RevBayes_remarks 09/08/2018 14:35 Archivo TXT 2 KB
RevStudio 07/08/2018 11:29 Aplicación 45.996 KB
zlib1.dll 01/08/2018 14:54 Extensión de la apl... 92 KB
hentos seleccionados 89,8 MB
RevBayes
File Edit Run
Script
Output
RevBayes version (1.0.9)
Build from development (f78150) on Sun, Aug 05, 2018 2:24:35 AM
Visit the website www.RevBayes.com for more information about RevBayes.
RevBayes is free software released under the GPL license, version 3. Type 'license()' for details.
To quit RevBayes type 'quit()' or 'q()'.
Command

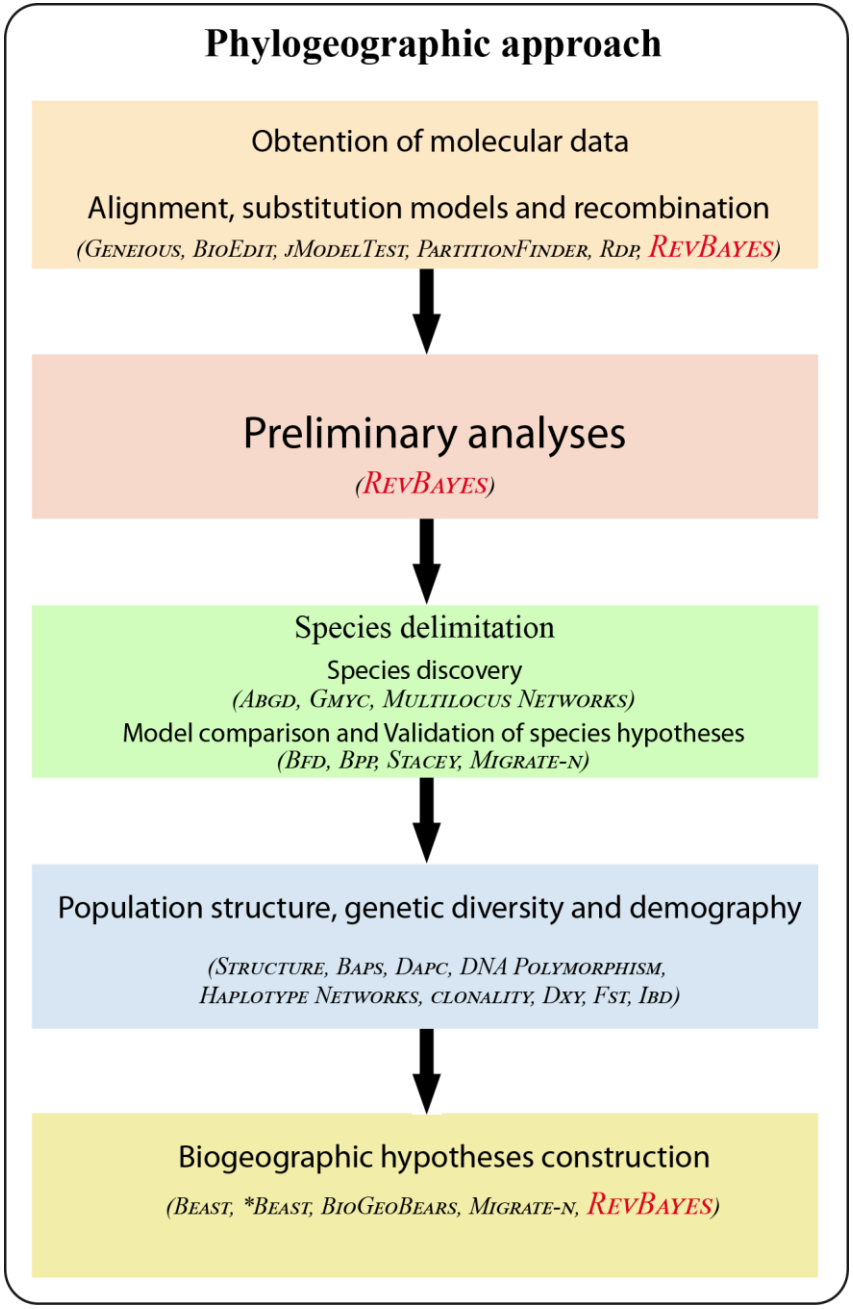


The data (alignment in nexus format) and the RevBayes script should be in the same directory together with the application “RevStudio”. Then, just click “source script”!!
An output folder will be automatically created containing the analysis results.



Introduction to RevBayes and MCMC
Model Selection and Testing
Standard tree inference
Complex hierarchical model for phylogenetic inference

Diversification Rate Estimation
Comparative methods
Biogeography



RevBayes offers an unique analytical framework into which many popular analyses in phylogeography can be conducted

Software Tutorials Workshops Jobs Developer

RevBayes Workshops

Throughout the year, the members of the RevBayes development team and our collaborators teach workshops on molecular evolution, phylogenetics, and Bayesian inference using RevBayes. Additionally, we have occasional hackathons which bring together developers to work on the software and methods for phylogenetic analysis.

Future Workshops

Date	Course Title	Location	Instructors
May 25, 2019	Bodega Applied Phylogenetics Workshop	Bodega Bay, California, USA	Brian Moore, Cécile Ané, John Huelsenbeck, Sebastian Höhna, Michael Landis, Mike May, Bruce Rannala, Bob Thomson, Peter Wainwright
December 11, 2018	Analysing Macroevolutionary Processes using RevBayes	Université de Montpellier, Montpellier, France	Sebastian Höhna, Rachel Warnock, Fabien Condamine, Thomas Couvreur

Past Workshops

Date	Course Title	Location	Instructors
October 5, 2018	Bayesian Phylogenetics in RevBayes	The Field Museum, Chicago, IL USA	Tracy Heath
August 2, 2018	MadPhylo: Madrid Workshop on Phylogenetics	Royal Botanical Garden, Madrid, Spain	Sebastian Höhna, John Huelsenbeck, Brian Moore, Fredrik Ronquist, Isabel

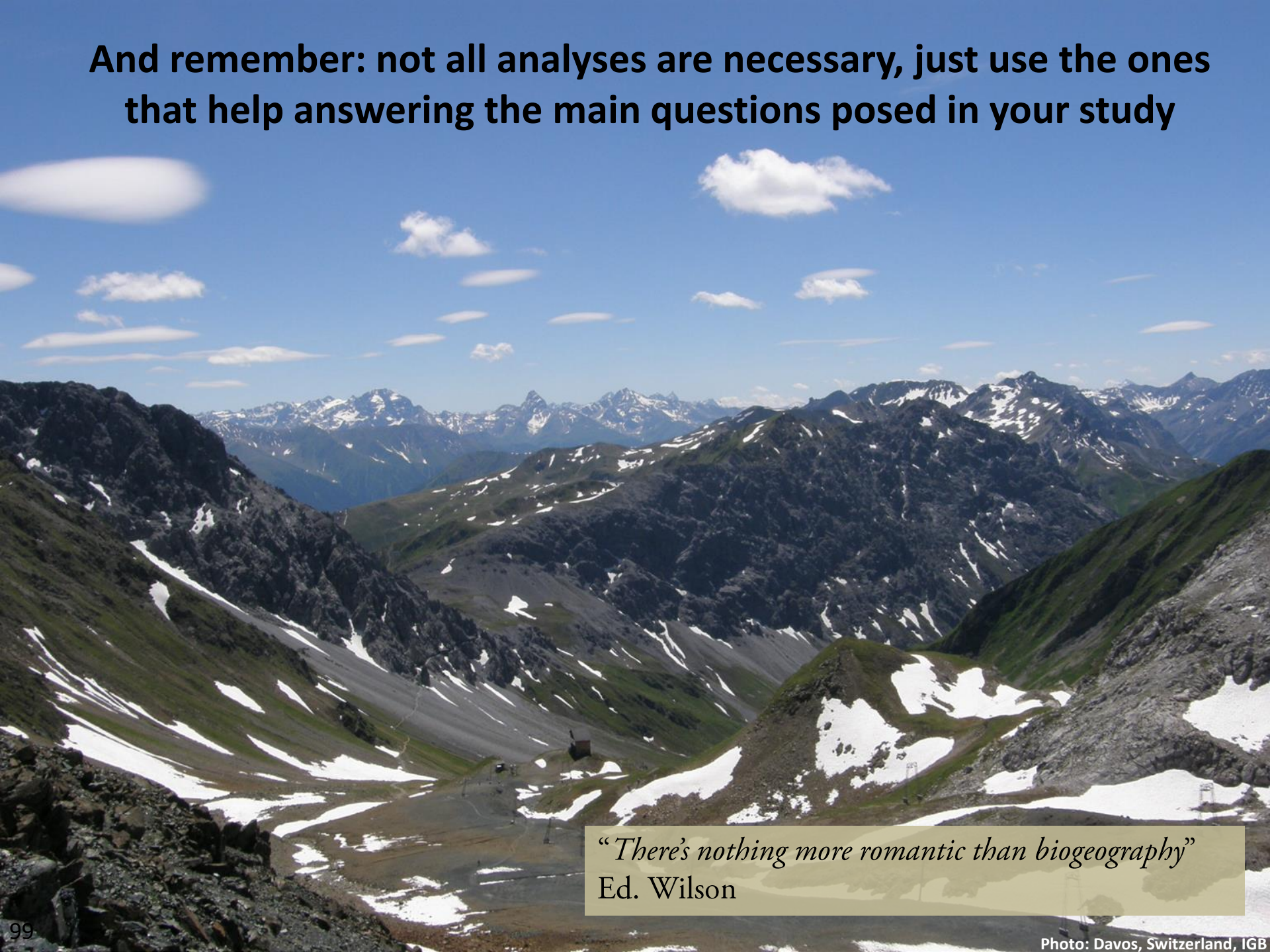
Google Scholar: 67 citations in 2 years

Major advantage: RevBayes constitutes a unique Bayesian framework that integrates analyses ranging from estimation of the best substitution model to biogeography and trait evolution.

In my opinion, **two major drawbacks**:

- "Proficiency" in statistical phylogenetics and modelling (if you are already proficient, the Rev language offers you the possibility to tune many parameters of any analysis)
- Analyses are quite time-consuming because the program is still not available in online platforms such as CIPRES

And remember: not all analyses are necessary, just use the ones that help answering the main questions posed in your study



"*There's nothing more romantic than biogeography*"
Ed. Wilson

Photo: Davos, Switzerland, IGB