



Faculty of Environment and Natural Resources  
Department of Biometry and Environmental System Analysis

---

# Exploring the ecological and evolutionary drivers of diversification of the alpine plant genus *Phyteuma* (Campanulaceae)

BACHELOR THESIS

in

ENVIRONMENTAL SCIENCES

by

**Laura-Marie Vecera**

Matriculation number 332 60 16

Freiburg, 17th December 2014

Examiner: Prof. Dr. Carsten F. Dormann  
Second examiner: Dr. Cristina Roquet Ruiz

## **Erklärung zur Bachelorarbeit**

Hiermit erkläre ich, dass ich die Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die aus fremden Quellen direkt oder indirekt übernommenen Gedanken als solche kenntlich gemacht habe. Die Bachelorarbeit habe ich bisher keinem anderen Prüfungsamt in gleicher oder vergleichbarer Form vorgelegt. Sie wurde bisher auch nicht veröffentlicht.

(Datum)

(Unterschrift)

## Abstract

On the evidence of current anthropogenic climate change it becomes increasingly important to understand how it could affect different species. If they are incapable of adapting by evolving innovative traits, they can only persist if they migrate to areas where the former environmental conditions still exist. Investigating the evolutionary history of certain genera can help us to gain a better understanding for these processes. The plant genus *Phyteuma* is a suitable study group for this purpose. It has been exposed to changing climatic conditions several times during the quaternary. Nevertheless, none of its species show traits that hint at an evolution of climatic niches. It rather seems like populations followed the alpine climate and new species evolved after geographic separation and consequently reproductive isolation. The aim of this study was to test these hypotheses to learn more about the evolutionary history of *Phyteuma*. After the reconstruction of the phylogeny, geographical and ecological attributes of sister species, who share a common ancestral history, were compared to those of non-sister species. Complete geographic separation and niche conservatism are considered as prerequisites of allopatric speciation. Thus, sister species should have a low range overlap and a high overlap of climatic niches compared to non-sister species. Climatic niches were estimated by climatic niche modelling. Furthermore, it was tested whether secondary contacts could have led to a higher range overlap between older sister species. As changing environmental conditions can be expected to divide populations randomly, the range asymmetry between species pairs should have no clear tendency. Projections of occurrence of *Phyteuma* for the Last Glacial Maximum and today were produced based on the climatic niche models to see whether the adaptation to colder climates could have led to a decrease of range size in the course of global warming.

The results mainly confirm the hypotheses. Range asymmetry between species pairs was similar for sister species and non-sister species. The high overlap of climatic niches between sister species compared to non-sister species indicates that niche conservatism was prevalent in this genus. The predicted decrease of range size confirms this result. On the contrary, sister species had a higher range overlap than non-sister species. This pattern could emerge from secondary contacts as older sister species had a higher range overlap than younger ones. *P. spicatum* and *P. gallicum* seem to have evolved under sympatric speciation. *P. pyrenaiucm* and *P. vagneri* have clearly undergone a long distance dispersal event. In general, the results of this study indicate that the alpine plant genus *Phyteuma* predominantly diversified through allopatric speciation. Changes to warmer climates and the lack of capability for adaptation likely forced populations to track their preferred climate. Apparently, secondary contacts occurred during colder periods and could explain the co-existence of formerly allopatric species. To finally confirm all of these results, further research is necessary.

## Zusammenfassung

In Anbetracht des anthropogenen Klimawandels wird es immer wichtiger zu verstehen, wie dieser sich auf verschiedene Arten auswirken könnte. Fehlt ihnen die Fähigkeit zur evolutionären Anpassung, können sie ihr Fortbestehen nur durch Abwanderung in Gebiete mit erträglicheren Umweltbedingungen sichern. Ein Blick in die evolutionäre Vergangenheit bestimmter Gattungen kann zu unserem Verständnis dieser Prozesse beitragen. Die Gattung *Phyteuma* (Teufelskrallen) eignet sich hierfür gut. Im Laufe des Quartäres war sie mehrmals wechselnden Klimabedingungen ausgesetzt. Doch keine ihrer Arten besitzt Merkmale, die auf eine Anpassung an diese Veränderungen hinweisen. Populationen scheinen stattdessen dem alpinen Klima gefolgt zu sein und neue Arten könnten daraufhin durch geographische Trennung und reproduktive Isolation entstanden sein. Ziel dieser Arbeit war es diese Hypothesen zu testen und so mehr über die evolutionäre Geschichte von *Phyteuma* zu lernen. Nach der Rekonstruktion des Stammbaumes dieser Gattung, wurden geographische und ökologische Eigenschaften von Geschwisterarten, welche eine gemeinsame Abstammungsgeschichte teilen, mit denen von Nicht-Geschwisterarten verglichen. Eine vollständige geographische Trennung und die Beibehaltung der klimatischen Nische einer Art wird als Voraussetzung von allopatrischer Artbildung betrachtet. Daher sollten Geschwisterarten eine geringere Überlappung ihrer Areale und eine höhere Überlappung ihrer klimatischen Nische als Nicht-Geschwisterarten aufweisen. Die klimatischen Nischen wurden anhand von Modellen beschrieben. Es wurde außerdem getestet, ob sekundäre Kontakte zu einer höheren Überlappung der Areale von älteren Geschwisterarten geführt haben könnten. Da neue klimatische Bedingungen das Areal einer Art meist zufällig zerteilen, sollte das Verhältnis der Größe der Areale zweier Arten keine klare Tendenz aufweisen. Um zu ermitteln, ob sich die Areale der Arten von *Phyteuma* aufgrund ihrer Anpassung an das alpine Klima im Zuge einer globalen Erwärmung verringert haben könnten, wurden ihre Areale mit Hilfe der Nischenmodelle für das letzte glaziale Maximum und die gegenwärtigen Bedingungen vorhergesagt und verglichen.

Die Ergebnisse bestätigten größtenteils die oben genannten Hypothesen. Das Verhältnis der Größe der Areale zweier Arten war für Geschwister und Nicht-Geschwisterarten ähnlich. Die hohe Überlappung der klimatischen Nische der Geschwisterarten im Vergleich zu Nicht-Geschwisterarten weist auf eine Beibehaltung der klimatischen Nische im Laufe der Evolution von *Phyteuma* hin. Die vorhergesagte Verringerung der Areale seit dem letzten glazialen Maximum bestätigt dieses Ergebnis. Allerdings zeigten Geschwisterarten von *Phyteuma* eine höhere Überlappung ihrer Areale als Nicht-Geschwisterarten. Dieses Muster könnte durch sekundäre Kontakte entstanden sein, da die Überlappung der Areale von Geschwisterarten mit ihrem Alter steigt. *P. spicatum* und *P. gallicum* scheinen durch sympatrische Artbildung entstanden zu sein. Der Auslöser für die Evolution von *P. pyrenaicum* und *P. vagneri* könnte die Verbreitung von Samen über längere Distanzen gewesen sein. Allgemein weisen die Ergebnisse dieser Studie auf eine vorwiegend allopatrische Artbildung bei der alpinen Pflanzengattung *Phyteuma* hin. Wechsel zu wärmeren klimatischen Bedingungen und fehlende Akklimatisierung zwangen Populationen vermutlich dazu, ihren bevorzugtem Klima zu folgen. Anscheinend gab es sekundäre Kontakte während kälterer Perioden, welche das Vorkommen von ehemals allopatrischen Arten in ähnlichen Verbreitungsgebieten erklären könnten. Um all diese Ergebnisse endgültig zu bestätigen, ist weitere Forschung notwendig.

# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Material and Methods</b>	<b>5</b>
2.1. Study group <i>Phyteuma</i> (Campanulaceae) . . . . .	5
2.2. Phylogeny . . . . .	5
2.3. Occurrence data . . . . .	7
2.4. Climatic niche modelling . . . . .	7
2.5. Niche overlap . . . . .	9
2.6. Range overlap and range asymmetry . . . . .	10
<b>3. Results</b>	<b>13</b>
3.1. Phylogeny . . . . .	13
3.2. Climatic niche modelling . . . . .	13
3.3. Niche overlap . . . . .	19
3.4. Range overlap and range asymmetry . . . . .	19
<b>4. Discussion</b>	<b>23</b>
4.1. Main findings and comparison to other studies . . . . .	23
4.2. Limitations . . . . .	24
<b>5. Conclusions and future perspectives</b>	<b>27</b>
<b>Appendix A. Information Occurrences <i>Phyteuma</i></b>	<b>34</b>
<b>Appendix B. Phylogeny</b>	<b>36</b>
<b>Appendix C. Climatic niche modeling</b>	<b>37</b>

## List of Figures

1.	Sketch of hypotheses . . . . .	3
2.	Pictures <i>Phyteuma</i> . . . . .	6
3.	Sketch response curves climatic variables . . . . .	10
4.	Best maximum likelihood tree without dropped species . . . . .	14
5.	Results Principal Component Analysis . . . . .	15
6.	TSS scores ensemble models . . . . .	16
7.	Habitat suitability Last Glacial Maximum and current climate . . . . .	18
8.	Violinplot of niche overlap . . . . .	19
9.	Maps of ranges of sister species . . . . .	20
10.	Violinplots of range overlap and range asymmetry . . . . .	21
11.	Regression of divergence time and range asymmetry . . . . .	22
12.	Violinplot of projected range size Last Glacial Maximum and current climate	22
A.1.	Maps of distribution of <i>Phyteuma</i> . . . . .	35
B.1.	Best maximum likelihood tree . . . . .	36
C.1.	Weighted Mean Probabilities versus Committee Averaging . . . . .	37

# List of Tables

1. Information species niche . . . . . 17

# 1. Introduction

As current anthropogenic climate change proceeds, it becomes increasingly important to understand the impact of shifting environmental conditions on species distributions (Bellard et al., 2012). The occurrence of a species is usually limited by its ecological niche, the abiotic and biotic variables that enable it to persist in nature (Hutchinson, 1957). Changing conditions of the environment of a species force it to react by migrating or adjusting its ecological niche through adaptation. Species are endangered by extinction if they cannot evolve fast enough and have limited dispersal or migration capabilities (Thuiller et al., 2008). An examination of species evolutionary past can reflect those processes and give us important information, e.g. for conservation planning (Thuiller et al., 2008; Boucher et al., 2011).

In theory, different speciation events are conceivable in the course of climate change (Paiva et al., 2011). Speciation is the rise of new species through evolution. Climate change can be a driver for evolution as it puts populations under selective pressure (Bird et al., 2012). Species that already possess genes that might become advantageous under new conditions will be able to adapt (Hoffmann and Sgrò, 2011). As this can lead to a reduced gene flow with the rest of the population, a prerequisite of reproductive isolation, the origin of new species within the range of the original population becomes possible (Paiva et al., 2011). This process has been described as sympatric speciation (Fig. 1(a)). There is an ongoing discussion about the exact definition of sympatric speciation and its causes but the consensus is that it is clearly distinguishable from other speciation modes as it describes speciation without any spatial separation of populations (Bird et al., 2012). It is expected to be driven by disruptive natural selection and consequently ecological divergence (Bird et al., 2012). The species emerging from this process can therefore differ in their phenology, mating/pollinator system or ploidy (Anacker and Strauss, 2014).

As climate change has a direct impact on the environment a species lives in (e.g. temperature, availability of water, weather extremes) it can also raise insuperable barriers for populations depending on their capability to adapt (Wiens and Graham, 2005) (Fig. 1(b)). If a species can evolve traits that enable it to endure the new conditions, it is going to persist across the barrier (Wiens and Graham, 2005). On the contrary, niche conservatism, where a species is not capable of evolving fast enough to adapt to its environment, making it impossible for it to cross the barrier, can lead to fragmented populations (Kozak and Wiens, 2006). Slight differences in the gene pool of the separated populations and distinct environmental conditions they are subject to could lead to divergent ecological evolution (Peterson, 1999). This has been defined as allopatric speciation, and has long been considered as the most important process for diversification (Turelli et al., 2001).

Intermediate forms of these speciation modes exist. If a small part of a population adapts to ecological conditions at the edge of its population of origin, the gene flow might be interrupted sufficiently to favour the rise of a new species (Gavrilets et al., 2000). This is called parapatric speciation. Also, if a species has high dispersal capabilities, e.g. light seeds that can be easily transported by wind, a so called "founder population" might be able to colonize new habitats adjacent to its parent population (peripatric speciation) (Barraclough and Volger, 2000).



Information about the overlap and similarity of ecological niches of extant species pairs and an investigation on how their ranges relate to each other can be informative to reconstruct the evolutionary history of a genus or a clade (Barracough and Vogler, 2000; Wiens and Graham, 2005; Fitzpatrick and Turelli, 2006). As phylogenies inferred from genetic data of extant species represent cladogenetic splits, they can be used to trace processes involved in diversification (Barracough and Vogler, 2000). Sister species share a common ancestral history and this can have a direct imprint on their ecological and geographical attributes. Therefore, the examination of sister species may give valuable insights into the process and geography of speciation (Barracough and Vogler, 2000; Anacker and Strauss, 2014).

Sympatric and allopatric speciation should imprint differently on the resulting geographical and ecological patterns because they are based on disparate theoretical bases.

“Budding speciation”, which summarizes all processes where a small population is separated from the bigger one by micro-scale processes (parapatric, peripatric, sympatric as well as catastrophic or centrifugal speciation) (Anacker and Strauss, 2014), is considered to result in a high overlap of geographic ranges of sister species as one population splits off within or close to the other. Furthermore, a high range asymmetry between the populations of the two new species is expected to result from these processes (Fitzpatrick and Turelli, 2006). Studies in the past have shown that sympatric speciation is based on an ecological differentiation of the sister species (Rundle and Nosil, 2005). Taking this into account, overlap of their ecological niches should be low.

Allopatric speciation, on the contrary, is primarily caused by geographic barriers (Turelli et al., 2001). Upon this it follows that the range overlap of two recently diverged allopatric sister species should be low or not existent. It can be assumed that the ranges get fragmented randomly and therefore range asymmetry of species pairs should not show any trend (Barracough and Vogler, 2000). Niche conservatism is suspected to lead to vicariance and allopatric speciation (Wiens and Graham, 2005). Thus, sister species should share common ecological niches, respectively, and have a high niche overlap (Kozak and Wiens, 2006).

Environmental changes, like climate change, are expected to play a major role as drivers of diversification. Different modes of speciation seem to exist and a main concern of current research has been to find a clear definition of these processes and to investigate examples in nature. Kozak and Wiens (2006) combined phylogenies, geographical ranges, information about climatic niches and genetic information about sister populations to proof allopatric speciation for *Desmognathus* and *Plethodon* salamanders. With a similar approach, Anacker and Strauss (2014) showed that for plant species in the Californian Floristic Province “budding speciation”, where a new species forms within or at the edge of its sister species, has been predominant. Similarly, Grossenbacher et al. (2014) observed “budding speciation” in North American Monkeyflowers (*Mimulus spp.*). Even if studies like these have been conducted increasingly in the past, it remains uncertain whether speciation modes can actually be clearly differentiated. Aguilée et al. (2012) showed that landscape dynamics, where allopatric and sympatric processes alternate constantly, could likewise play an important role in diversification. Ranges of extant species could have changed over time and species which evolved in vicariance might share a common range today due to secondary contact (Barracough and Vo-

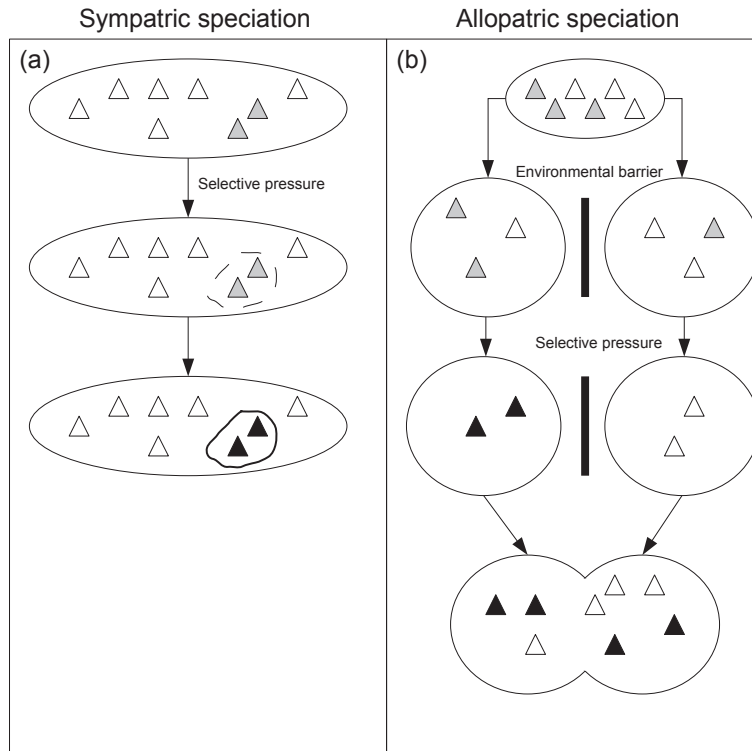


Figure 1: Sketch of Hypotheses: (a) Sympatric speciation: A population of the same species with phenotypic plasticity (open triangles and gray) exists within the same habitat. Climate change exerts selective pressure on the population. Some individuals, which were already pre-adapted to the new climatic conditions, evolve away from the sister population and might over time form their own species (black triangles) within the range of the sister species. This process can be identified by a high range overlap, a high range asymmetry and a low overlap of climatic niches between sister species. (b) Allopatric speciation: A population of the same species with phenotypic plasticity (open triangles and gray triangles) gets separated by an environmental barrier formed by climatic changes. Both sub-populations differ slightly in their gene pool and as they are reproductively isolated, mutation, genetic drift and selective pressures will on the long run lead to divergence. This process can be identified by a low range overlap, a random range asymmetry and a high overlap of climatic niches between sister species. If another climatic change takes place, the populations of the two species might mix up again (secondary contact). Range overlap can therefore increase with the time since species diverged.

gler, 2000). Geographical patterns that existed just after speciation can get blurred due to this (Fig. 1(b)). To avoid forming wrong conclusions from current species ranges, range overlap dependent on time of divergence of a pair of species can be taken into account (Barraclough and Vogler, 2000). Distantly related species, which diverged a long time ago in allopatric

conditions are expected to have a higher range overlap due to secondary contacts compared to young species pairs.

In this study, I am going to apply the approaches pointed out above to explore the evolutionary history of the plant genus *Phyteuma* (Campanulaceae). I hope to add a piece of the puzzle to our understanding of speciation in context with climatic change. *Phyteuma* is an alpine plant genus that is endemic to Europe and contains certain species also endemic to particular mountain ranges (e.g. *P. pyrenaicum* to the Pyrenees or *P. tetramerum* to the Carpathians) (Schulz, 1904; Schneeweiß et al., 2013). Schulz (1904) started to investigate this plant genus and published a whole monograph, trying to reveal its evolution by interpreting distributional patterns of *Phyteuma* and its phenomenological appearance. But new possibilities in genetics, informatics and new methods in evolutionary biology can now help us to gain a much more profound insight into the evolutionary history of this plant genus. Its phylogeny, recently published by Schneeweis et al. (2013), appears to be very balanced. Thus, no sudden increases of diversification rate can be expected and as current species do not show any obvious trait that could have led to ecologically divergent adaptation, the diversification of *Phyteuma* seems to be imprinted by geographical events and the conservation of climatic niches (in contrary to e.g. the plant genus *Androsace*, which shows an increase of diversification in two clades connected with the evolution of the cushion life form (Boucher et al., 2011)). As Schneeweis et al. (2013) date the first divergence events of *Phyteuma* back to about 3 million years ago, just when the first glaciation events of the Quaternary began, the diversification of this genus seems to be connected to climatic changes. Looking at the distribution of extant species, which appear to be predominantly alpine, it seems plausible to assume that the ancestors of *Phyteuma* tracked suitable climates during warmer periods. This could have led to a fragmentation of species populations and to a decrease of range size. Allopatric speciation might then have been triggered by reduced gene flow and reproductive isolation. This process could have repeated several times as glacial periods alternated with interglacial periods during the Quaternary. It is highly possible that secondary contact of species occurred during periods of glaciation when *Phyteuma* found more favourable conditions and its species were able to spread out again.

I aim to answer the following questions in this thesis: (1) Can climate change and niche conservatism be expected to be the main drivers of diversification in *Phyteuma*? (2) Did the ranges of species change in the course of climatic changes? (3) How much have post-speciation range changes and secondary contact contributed to the diversification of *Phyteuma*?

## 2. Material and Methods

### 2.1. Study group *Phyteuma* (Campanulaceae)

All species of *Phyteuma* show flowers in a dense, spikelet-like or head-shaped, terminal inflorescence surrounded by bracts to attract insects (Lauber and Wagner, 2007). The corolla is narrow tubed and the five corolla lobes are grown together on the base and fused at the tips (Lauber and Wagner, 2007). *Phyteuma* can grow up to 100 cm tall (Lauber and Wagner, 2007). Its colours range from purple to pale blue and white (Fig. 2). Its closest relative, the plant genus *Physoplexis*, lives in the Alps and is morphologically similar to *Phyteuma* (Schneeweiss et al., 2013). *Phyteuma* is pollinated by insects and dispersed by small seeds (Schulz, 1904). All species are perennials (Schulz, 1904). It is endemic to Europe and the species mainly inhabit mountain ranges of Central Europe (e.g. the Alps, the Pyrenees, the Carpathians, the Massif Central). Some species (*P. gallicum* (Fig. 2(b)), *P. pyrenaicum* (Fig. 2(c)), *P. serratum*, *P. tetramerum* and *P. vagneri* (Fig. 2(d))) are endemic to the mountain ranges they live on (see Fig. A.1(d)). *P. spicatum* (Fig. 2(a)), *P. orbiculare* and *P. nigrum* are widespread species also occurring in the lowlands of Europe and *P. charmelii* is the only species that occurs outside of Europe, in northern Africa (Schulz, 1904; Quézel, 1954). Species can be found on siliceous or calcareous bedrocks (Schulz, 1904), in forests and meadows (Landolt, 2010).

### 2.2. Phylogeny<sup>1</sup>

Recently, Schneeweiss et al. (2013) published a complete phylogeny of the plant genus *Phyteuma* (Campanulaceae) based on molecular data. For this study sequences of the combined alignment produced by Schneeweiss et al. (2013) were downloaded from TreeBASE (study ID 12627). They used plant material of 24 *Phyteuma* species (including two subspecies each of *P. globulariifolium* and *P. scheuchzeri*) and of 106 individuals in total. Three DNA regions were included in the combined alignment: the nuclear region ITS and the chloroplastic regions *trnL-F* and *matK*. Phylogenetic inference was conducted with maximum likelihood (ML) as implemented in *RAxML* v.8 (Stamatakis, 2014). 100 independent searches were carried out to retrieve 100 ML trees. Node support was assessed by bootstrap analysis with 1000 replicates. When species were monophyletic, all individuals except one per species were pruned. In the case of the species *P. spicatum* and *P. ovatum*, which were interbred, the two species were considered as a single evolutionary entity (*P. spicatum* sensu lato) and all individuals were removed except for one to represent this group. The same procedure was applied with *P. charmelii* and *P. villarsii*, which were considered as *P. charmelii* s.l. . Phylogenetic trees were dated with *r8s* (Sanderson, 2003), calibrating the node divergence between *Asyneuma* and *Phyteuma* with a minimum age of 12.4 and a maximum age of 8.8 based on the 95% confidence interval for the divergence between these two genera as reported in Olesen et al. (2012).

---

<sup>1</sup>All analyses described in this section have been conducted by Cristina Roquet

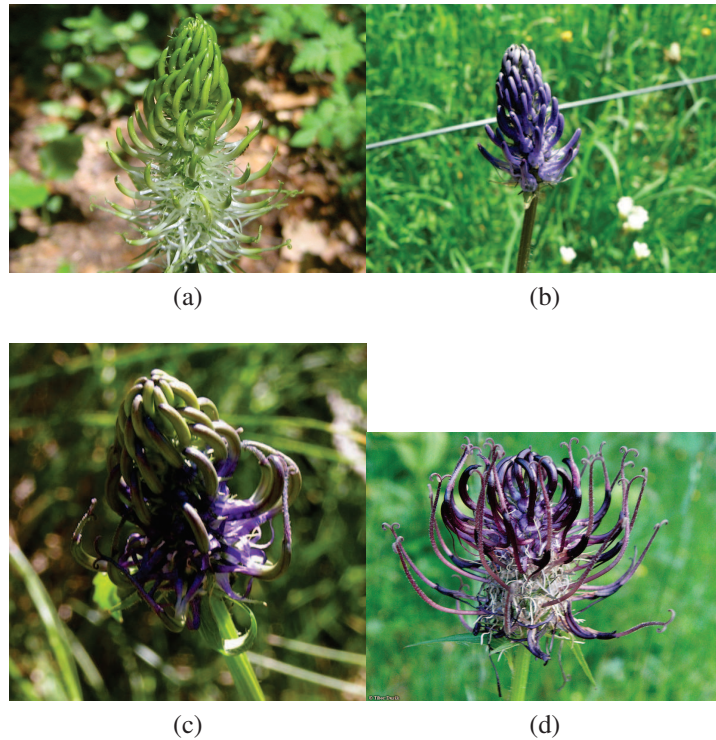


Figure 2: Pictures of selected sister species of *Phyteuma*  
 (a) *P. spicatum* ([www.alpenflora.ch/phyteuma-spicatum/](http://www.alpenflora.ch/phyteuma-spicatum/))  
 (b) *P. gallicum* (Mathieu Menand,  
[www.tela-botanica.org/bdtfx-nn-49211-synthese](http://www.tela-botanica.org/bdtfx-nn-49211-synthese))  
 (c) *P. pyrenaicum* (Alain Bigou,  
<http://www.tela-botanica.org/bdtfx-nn-49262-illustrations>)  
 (d) *P. vagneri* (Tibor Dusik,  
[www.carpathianbasinspecies.eu/2\\_Vascular\\_Plants%28Tracheophyta%29/  
 Species/Phyteuma\\_vagneri/Phyteuma%20vagneri\\_02\\_1\\_logo\\_0.html](http://www.carpathianbasinspecies.eu/2_Vascular_Plants%28Tracheophyta%29/Species/Phyteuma_vagneri/Phyteuma%20vagneri_02_1_logo_0.html))

To assess and quantify the degree of tree incongruence between the 100 ML trees inferred, the “internode certainty all” (ICA) and the “tree certainty all” (TCA) indices (Salichos et al. 2014) were computed as implemented in *RAxML v8*. The ICA calculates the degree of certainty for a given internode by considering the frequency of the bipartition defined by the internode (internal branch) in a given set of trees in conjunction with that of all conflicting bipartitions in the same underlying tree set. Internode-certainty values near zero indicate the presence of an almost equally supported bipartition that conflicts with the inferred internode, whereas values close to one indicate the absence of conflict. The TCA is the sum of ICA values across all internodes of a phylogeny.

Sister species have been defined here as all species, which directly share a common ancestor that they do not share with any other species of the tree.



### 2.3. Occurrence data

Georeferenced species occurrence data for extant species were collected using different databases, i.e. the open Global Biodiversity Information Facility database (GBIF, [www.gbif.org](http://www.gbif.org), last accessed 21.03.2014) and various local and personal sources (Conservatoire Botanique National Alpin (CBNA, [www.cbn-alpin.fr](http://www.cbn-alpin.fr)), Jonathan Lenoir (Jules Verne University of Picardie ([www.jonathanlenoir.wordpress.com/datasets/](http://www.jonathanlenoir.wordpress.com/datasets/))), Hurdu Bogdan-Iuliu (Institute of Biological Research Cluj (Romania)([www.icbcluj.ro](http://www.icbcluj.ro))) and Niklaus E. Zimmermann (Swiss Federal Research Institute WSL ([www.wsl.ch/info/mitarbeitende/zimmerma/index\\_EN](http://www.wsl.ch/info/mitarbeitende/zimmerma/index_EN))). As described in subsection 2.2 all occurrences of *P. spicatum* and *P. ovatum* as well as *P. villarsii* and *P. charmelli* were merged. Furthermore, all points of a resolution lower than 0.01 arc-degree and points with erroneous coordinates (e.g. situated in the sea) as well as duplicated entries were removed so that a dataset of 33073 occurrence points remained (Fig. A.1). 5 occurrences are the minimum to have sufficient knowledge about the distribution of a species to be able to formulate a decent ecological niche model (Hernandez et al., 2006; Boucher et al., 2011). No occurrences were found for *P. pseudorbiculare* and only 4 occurrences could be obtained for *P. confusum*, therefore they were excluded from all further analyses. On average 1574 points of occurrence could be obtained for each species ranging from 8 occurrences for *P. persicifolium* to 13909 occurrences for the widespread species *P. spicatum* (table 1). Only for three species were less than 10 occurrences found (*P. gallicum*, *P. persicifolium* and *P. confusum*).

### 2.4. Climatic niche modelling

Lately, ecological niche models (ENMs) and climatic niche models (CNMs) in particular have been used a lot in studies of niche evolution (Kozak et al., 2008; Evans et al., 2009; Anacker and Strauss, 2014). CNMs are statistically preferable to raw climatic data as they can account for habitat availability in the area of interest (Carsten Dormann, pers. comm.). Occurrence data and climatic variables are being combined to obtain CNMs (see Guisan and Thuiller (2005) for a review). The aim is to find the best model to statistically relate the geographical distribution of species to the climatic conditions in which they occur (Guisan and Zimmermann, 2000). One of the outputs of climatic niche modelling is a map of habitat suitability based on the climatic variables that have been used to formulate the model.

To get information about the climatic preferences *Phyteuma*, data for 19 bioclimatic variables were extracted from the Worldclim database on a high resolution of 2.5 arc-minutes ([www.worldclim.org/bioclim](http://www.worldclim.org/bioclim))<sup>2</sup>. They were generated through interpolation of average monthly climate data from weather stations on a 30 arc-second resolution grid (Hijmans et al. 2005). The rasters of the climatic variables were cropped to an European scale. A Principal Component Analysis (PCA) was carried out using the *dudi.pca* function (package *ade4*). All 19 climatic variables based on the occurrence localities of all species were used to detect highly correlated variables within the species' climatic ranges and to reduce the number of variables to be included in the climatic niche modelling. This is important, as an increasing

<sup>2</sup> This and all further analyses were conducted using the R Software Version 3.1.2 (R Core Team, 2014)

number of parameters used for modelling can increase the fit of the model but will also increase the incorporation of errors (Diniz Filho et al., 2010). The scale option was set to TRUE so that the PCA was carried out on the correlation matrix to ensure that all variables have similar ranges (Dormann, 2013).

The package *BIOMOD2* (Version 3.1-48) was used to model climatic niches (see Thuiller et al. (2009) for a documentation of the first version of the *BIOMOD* package and Zimmerman (2009) for an opinion on *BIOMOD*). This program enables the user to compare outputs of a large number of different models (e.g. Generalized Linear and Additive Models (GLM and GAM), Boosted Regression Trees (GBM) and Random Forest (RT)) to find the model that can capture the relation between occurrences and climatic variables best. Furthermore, the user can extract ensemble models from different initial conditions, model classes and model parameters. In their review about ensemble forecasting, Araújo and New (2007), remark that ensemble models are usually qualitatively better than using only a single combination of data as they can obtain a lower mean error by combining the independent information contained in each separate model run. In this way, uncertainty can be reduced (Marmion et al., 2009).

As our dataset contained only presence data for extant species and as *BIOMOD2* requires presence-absence data for its calculations, 3 sets of pseudo-absences were produced for each species using the random sampling technique implemented in the *Biomod\_FormatingData* function. Pseudo-absences were sampled randomly from the whole environmental raster given to the function. To account for the number of occurrences available for a species and to keep up statistical power, the number of pseudo-absences produced was put up to 500 for species with less than 100 occurrences. For species with more than 100 occurrences the number of pseudo-absences produced was the same as the number of occurrences.

Because I wanted to use the same model for all of my species, I needed one that would potentially be flexible enough to represent all relationships between species distributions and bioclimatic variables. I decided to use GAM only, which is similar to GLMs but uses smoothing functions to link explanatory and response variables (for a detailed explanation, see Zuur et al. (2009)). This makes GAM less restricted for data fitting compared to GLM and especially suitable to describe non-linear patterns (Zuur et al., 2009). Response curves of species to ecological variables are mostly not linear but show distortions as they reflect the realized niche of a species which is often influenced by biotic interactions (Austin, 2002). This is why GAM is an especially attractive model for our purpose. Within the *BIOMOD\_ModelingOptions* function the algorithm *GAM\_mgcv* and its default settings were chosen to use the package *mgcv* to fit the GAM model. To enable an evaluation of the model being produced by the algorithm and due to the lack of independent data for evaluation, the dataset was split into 80% data for calibration and 20% data for evaluation. For each set of pseudo-absences the model was run 4 times to get a measure for uncertainty of the model. Thus at the end of the modelling process 12 models were obtained (3 sets of pseudo-absences and 4 sets of cross-validation). By default

the algorithm returns the evaluation metric True Skill Statistic (TSS)<sup>3</sup>. Ensemble modelling by weighted mean of probabilities was used to merge all of these models with a TSS score higher than 0.7<sup>4</sup>.

To get the habitat suitability of each pixel of the provided climatic rasters, projections of potential species distributions were generated from the models. Ensemble forecasts could be obtained including these projections and the ensemble models.

To obtain information about the preferences of species concerning selected climatic variables, an approach first formulated by Evans et al. (2009) was followed to extract meaningful response curves (Fig. 3). First, each climatic variable was reclassified into 100 discrete and equally sized classes. Afterwards based on the reclassified climatic raster, all habitat suitabilities per class were extracted and summed up. This value was then divided by the sum of all habitat suitabilities. Alternatively, histograms of habitat suitability per class could have been produced by just using the mean of habitat suitability per climatic class (sum of habitat suitability divided by number of cells per class). The first approach has been chosen because it keeps the information about the size of each climatic class in the area of interest.

## 2.5. Niche overlap

To calculate niche overlap, Schoener's D was used to quantify the similarity of climatic niches of different species (Schoener, 1968; Warren et al., 2008). It is implemented in the *niche.overlap* function of the package *phyloclim*. This metric has been well known for decades and has been recently applied in several studies with a focus on the evolutionary history of species (Warren et al., 2008; Evans et al., 2009; Anacker and Strauss, 2014). The variables  $p_X$  and  $p_Y$  are the probability of occurrence of species X (Y resp.) in a cell of the raster obtained from the CNM (Warren et al., 2008). Schoener's D is being calculated as follows:

$$D(p_X, p_Y) = 1 - \frac{1}{2} \sum_i |p_X - p_Y|$$

---

<sup>3</sup>The TSS is calculated as

$$\text{TSS} = \frac{\text{hits}}{\text{hits} + \text{misses}} - \frac{\text{false alarms}}{\text{false alarms} + \text{correct negatives}}$$

<sup>4</sup>There are several algorithms available in *biomod2* to carry out the ensemble modelling. Marmion et al. (2009) identified average functions (such as weighted mean) as the best consensus methods to increase accuracy of species distribution forecasts. To check whether there is a risk that different strategies, from which ensembles can be obtained, return extremely divergent results, the TSS scores of ensemble models produced by weighted mean of probabilities and committee averaging were compared to each other. Weighted mean of probabilities only was chosen for the ensemble modelling as it showed a positive linear relationship to committee averaging (Fig. C.1)



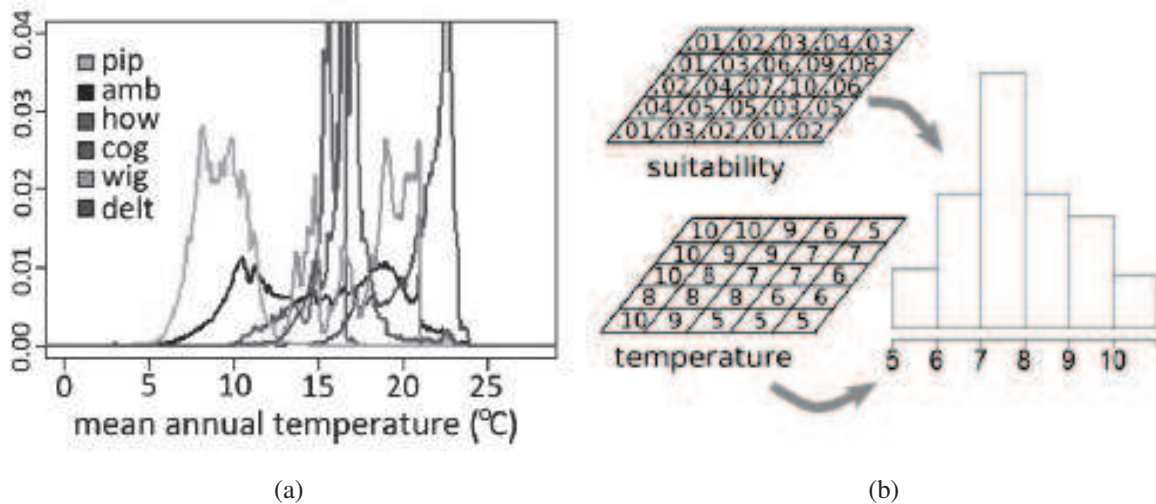


Figure 3: Figure adapted from Evans et al. (2009) to illustrate the process of obtaining optimal values for each climatic variable from the climatic niche modelling. (a) The graph shows curves of predicted niche occupancy of species Evans et al. (2009) used for their study. (b) Schematisation of how to obtain these histograms from the habitat suitability predicted by CNMs (see subsection 2.4).

Warren et al. (2008) state that it is advantageous to use probability distributions based on geographical instead of environmental space as they exclude all environmental space which can not actually be found in the area of distribution. This avoids over-prediction and therefore improves the calculation of niche overlap, too (Peterson, 2011).

To choose the right significance test for the data, first the Shapiro-Wilk normality test as implied in the *shapiro.test* function was conducted. As both of the groups were not normally distributed, a one-tailed Wilcoxon rank sum test (function *wilcox.test*) was used to see whether niche overlap was higher for sister species compared to non-sister species.

## 2.6. Range overlap and range asymmetry

The range overlap of species pairs was calculated following the example of Anacker and Strauss (2014). The range of each species was estimated by creating a 10 km buffer around each occurrence point using the *gBuffer* function (package *rgeos*) and by calculating the area of all of those polygons.

To calculate the range overlap of two species, the method formulated by Barraclough and Vogler (2000) was used:

$$\text{Range overlap} = \frac{\text{Area of overlap}}{\text{Range size of species with smaller range}}$$

This returns a value between 0 (no range overlap) and 1 (complete range overlap) and gives information about the degree by which the smaller ranged species is embedded into the range of the larger ranged species. To avoid overestimation of the range shared by both species, the intersect of both ranges was calculated using the *joinPolys* function (package *PBSmapping*) (Anacker and Strauss, 2014). Range overlap was calculated for each species pair, 190 pairs altogether.

Range asymmetry was calculated following Anacker and Strauss (2014):

$$\text{Range asymmetry} = \frac{\text{Area of larger ranged species}}{\text{Area of smaller ranged species}}$$

To test for significance between sister and non-sister species the same statistical procedure was conducted as explained in subsection 2.5. Similar to niche overlap, range overlap and range asymmetry were not normally distributed for at least one of the groups, so the Wilcoxon rank sum test was used to compare sister and non-sister species. As sister species were expected to have a higher range overlap than non-sister species, a one-tailed test was conducted. For range asymmetry, no strong a priori assumptions were made concerning the relation of range asymmetry of sister species to non-sister species, so a two-tailed test was carried out.

Furthermore, it was explored whether a relationship between the age of a species pair and its range overlap exists. The phylogenetic variance-covariance matrix (as implemented in the *vcv.phylo* function, package *ape*) was used to obtain the time shared by two species before divergence. This measure was translated into age by subtracting the time shared before divergence from the root age of 6.44 Million years. A linear regression between age and niche range overlap was conducted.

To test whether range size of species of *Phyteuma* changed between periods of colder and warmer climates, the climatic niche determined by the model calibrated for current climatic conditions has been projected on the climatic conditions of the Last Glacial maximum (LGM) (22,000 years ago). During the LGM, the mean annual temperature in Europe was clearly lower than today (about -6.4°C compared to 7.1°C today). The climatic layers were obtained from the Worldclim database on a spatial resolution of 2.5 arc-minutes ([www.worldclim.org/bioclim](http://www.worldclim.org/bioclim) (last accessed 27.10.2014)). The *BIOMOD\_EnsembleForecasting* function produces habitat suitability layers as well as binary layers where presences/absences are predicted for all cells. They are being extracted from the habitat suitability layers by imposing a threshold above which a presence is assigned to the cell. The threshold retained is the one

leading to the best ensemble model evaluation. The number of presences in the binary layers was summed and considered as the projected range size. To ensure comparability, projected ranges were also used for current climate. A one-tailed Wilcoxon signed rank test was used to infer whether range size was bigger in the past than today.

## 3. Results

### 3.1. Phylogeny

The likelihoods of the 100 ML trees obtained varied only slightly between - 13598.67 and - 13605.9 for the best ML tree. The bootstrap analysis showed a very high robustness of most of the nodes (85% of nodes obtained a node support >70). All nodes obtained an ICA value equal to one, and the relative TCA (TCA normalized by the maximum possible TCA scores for a fully bifurcating tree) obtained here for the set of 100 ML trees is 0.97. Thus, it can be assumed that the 100 ML trees are highly similar (which is congruent with the high bootstrap values obtained) and the subsequent analyses will therefore be performed with the best ML tree (Fig. B.1). As there could be no distribution data obtained for *P. pseudorbiculare* and *P. confusum* the tips of these two species were dropped from the tree so that 20 species remained (Fig. 4). Six sister species pairs exist within *Phyteuma*:

- *P. cordatum* and *P. michelii*
- *P. globulariifolium* and *P. hemisphaericum*
- *P. orbiculare* and *P. sieberi*
- *P. persicifolium* and *P. betonicifolium*
- *P. spicatum* and *P. gallicum*
- *P. vagneri* and *P. pyrenaicum*

Originally, *P. confusum* and *P. serratum* formed a sister species and *P. globulariifolium* and *P. hemisphaericum* a polytomy (Fig. B.1). But as *P. confusum* had to be dropped from the tree due to the lack of distribution data, I decided to consider *P. globulariifolium* and *P. hemisphaericum* as sister species even though they now built a polytomy with *P. serratum* (Fig. 4).

### 3.2. Climatic niche modelling

The first five axes of the PCA explained 96.35% of the variation in the data. Only two were included for the choice of climatic variables because they already explained 78.32% of the variation. Taking into account how much each climatic variable loaded onto those two axes and using the knowledge that extreme climatic conditions often limit species distributions, the following climatic variables were picked to carry out the climatic niche modelling: Annual Mean Temperature (Bio 1), Mean Temperature of the Warmest Quarter (Bio 10), Annual Precipitation (Bio 12), Precipitation Seasonality (Bio 15) and Precipitation of the Wettest Quarter (Bio 16) (Fig. 5). As *Phyteuma* is assumed to be an alpine plant genus, it can be expected that the extremes of precipitation and in particular of temperature (e.g. mean temperature of warmest quarter) are limiting factors to the occurrence of *Phyteuma*.

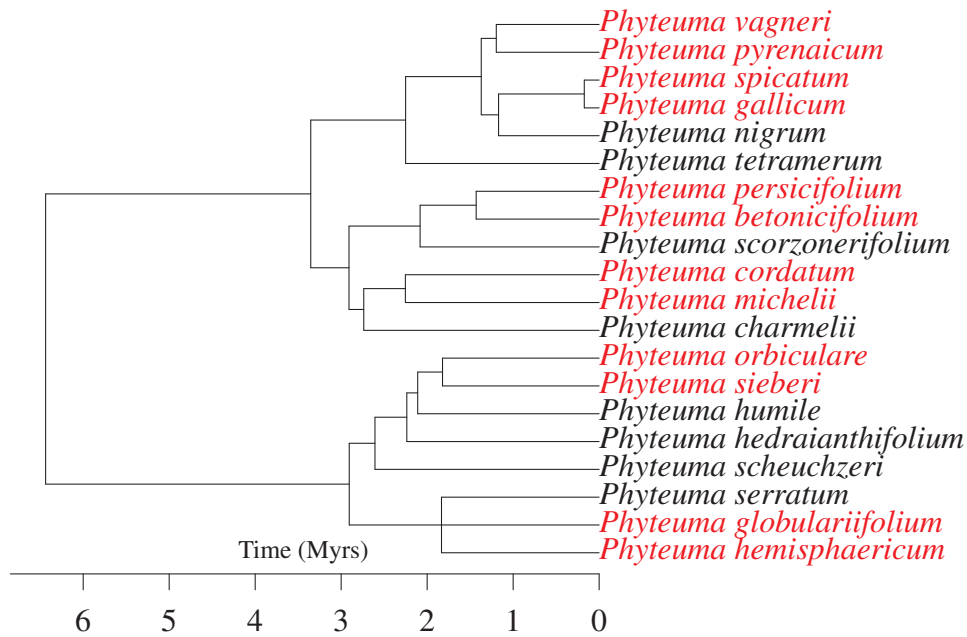


Figure 4: Best maximum likelihood tree obtained by the phylogenetic inference. The tips of the species *P. pseudorbiculare* and *P. confusum* have been dropped. The branches are dated and the scale bar represents time in Million years. Sister species are marked in red.

In general, the climatic niche modelling predicts the species of *Phyteuma* to occur in a relatively cold and humid climate with a mean annual temperature of  $7.3 \pm 2.6^\circ\text{C}$  and an annual precipitation of  $927.55 \pm 257.42$  mm (table 1). The widespread species *P. orbiculare* and *P. spicatum* seem to prefer a warmer and drier climate compared to the other species (mean annual temperature of  $10^\circ\text{C}$  ( $9.5^\circ\text{C}$ , resp.) and annual precipitation of 784.13 mm (both species)). *P. gallicum*, an endemic to the Massife Central in France, occurs in a warmer and drier climate (mean annual temperature of  $9.5^\circ\text{C}$  and annual precipitation of 701.39 mm). The species occurring in the coldest climates are *P. humile* and *P. hedraianthifolium* which can stand mean annual temperatures of  $0.1^\circ\text{C}$  ( $1.9^\circ\text{C}$ , resp.). They are also the species who populate areas with the highest annual precipitation of 1611.53 mm (1280.57 mm, resp.) compared to the rest of the genus.

The ensemble models obtained by the climatic niche models generally obtained quite high TSS scores (Fig.6). That means they were able to predict the evaluation occurrence points with high accuracy. Looking at the habitat suitability of sister species for the current climate

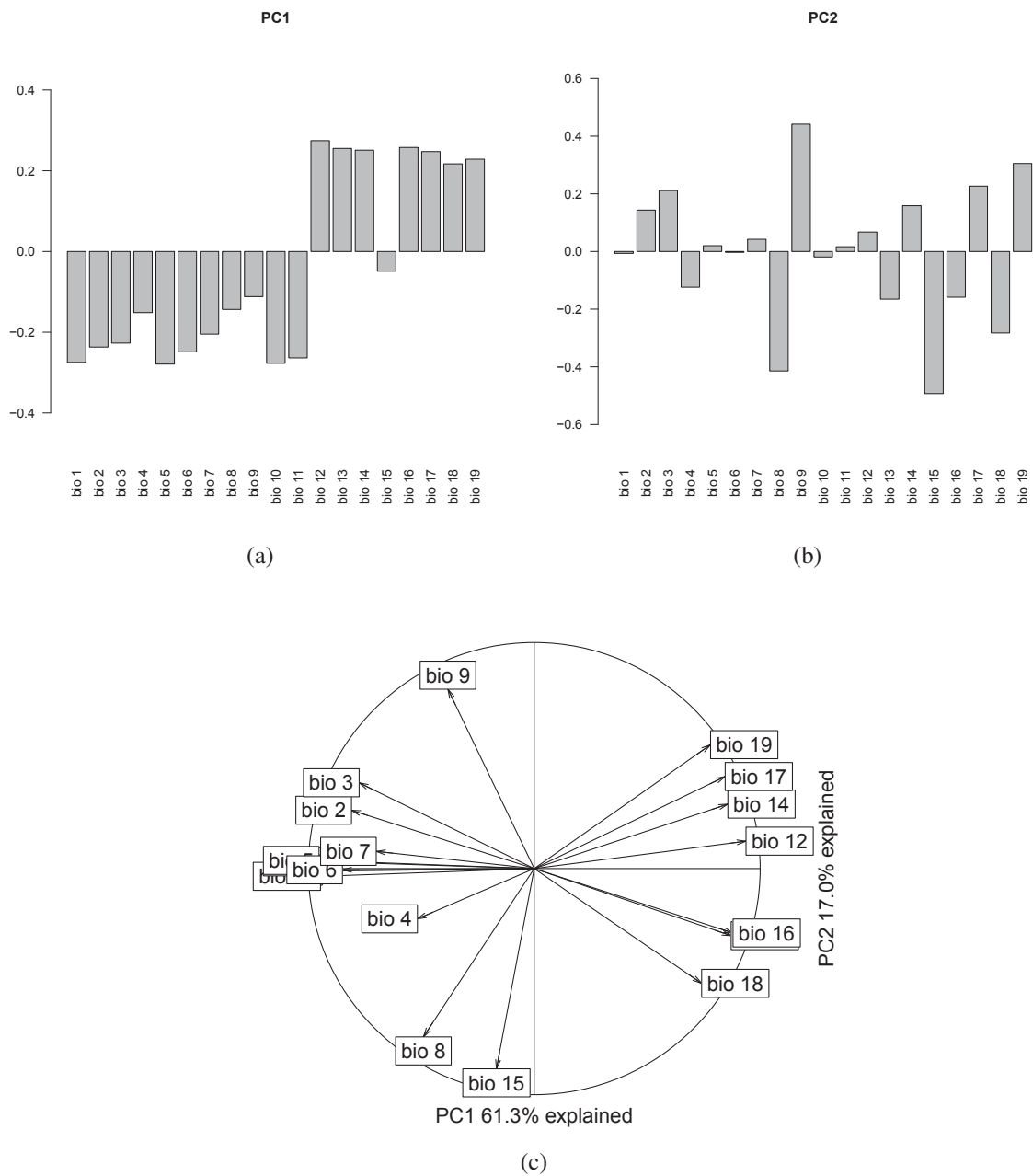


Figure 5: (a) Barplot of the loading of each climatic variable on the first principal component (PC1), (b) barplot of the loading of each climatic variable on the second principal component (PC2), (c) biplot of PC1 and PC2 and the loadings. The percentages give information of the amount of variance explained by each principal component.

it appears that the models assign a high suitability to areas where it is known that the species do not at all or only very rarely occur (e.g. Scandinavia and Great Britain) (Fig. 7). Included in the projections for the past climate areas predicted to be highly suitable are areas that are known

to have been widely covered by glaciers during the LGM. It is improbable that *Phyteuma* managed to survive in those areas.

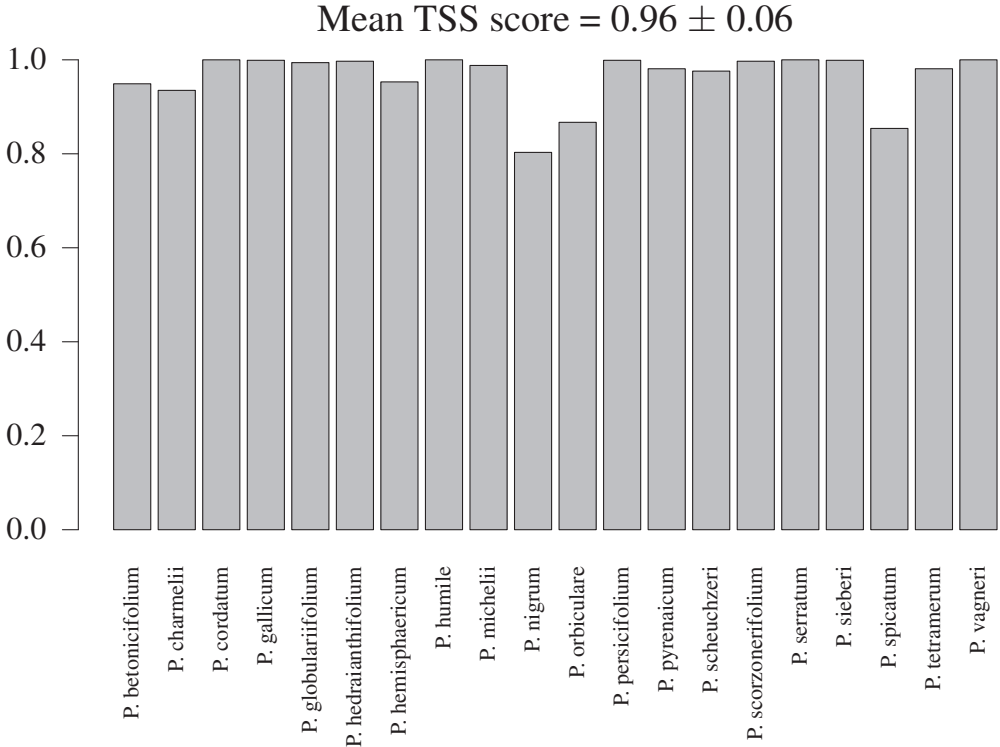


Figure 6: Barplot of the TSS scores for the GAM ensemble model for each species obtained by weighted mean of probabilities.

Table 1: Mean annual temperature (°C) (MAT), annual precipitation (mm) (AP), mean temperature of the warmest quarter (°C) (MTWQ), precipitation seasonality (Coefficient in Variation) (PS) and precipitation of the wettest quarter (mm) (PWQ) withdrawn from modelled climatic niches using the method by Evans et al. (2009) (see subsection 2.4).

<b>Species</b>	<b>Occurrences</b>	<b>MAT (°C)</b>	<b>AP (mm)</b>	<b>MTWQ (°C)</b>	<b>PS (Coefficient in Variation)</b>	<b>PWQ (mm)</b>
<i>P. betonicifolium</i>	3695	6.6	866.87	15.4	24.5	295.45
<i>P. charmelii</i>	318	8.9	591.07	16	19.5	276.23
<i>P. cordatum</i>	13	7.7	1004.77	13.6	12.5	295.45
<i>P. gallicum</i>	9	9.5	701.39	17.7	11.5	218.57
<i>P. globulariifolium</i>	1372	7.2	839.29	17.7	26.5	324.28
<i>P. hedraianthifolium</i>	18	1.9	1280.57	9.5	25.5	429.99
<i>P. hemisphaericum</i>	2559	8	1032.35	13.6	24.5	324.28
<i>P. humile</i>	25	-0.1	1611.53	8.9	20.5	449.2
<i>P. michelii</i>	596	7.7	922.03	13.6	22.5	285.84
<i>P. nigrum</i>	375	8.9	618.65	16	28.5	218.57
<i>P. orbiculare</i>	9526	10	784.13	17	11.5	276.23
<i>P. persicifolium</i>	8	8	1115.09	15.4	27.5	372.33
<i>P. pyrenaicum</i>	22	7.2	811.71	15.4	15.5	266.62
<i>P. scheuchzeri</i>	355	8	1225.41	16	24.5	391.55
<i>P. scorzonerifolium</i>	77	6.6	949.61	15.4	22.5	295.45
<i>P. serratum</i>	10	6.9	1004.77	18.4	36.5	362.72
<i>P. sieberi</i>	24	8.3	1115.09	17.1	25.5	391.55
<i>P. spicatum</i>	13909	9.5	784.13	16	11.5	218.57
<i>P. tetramerum</i>	78	8.3	646.23	16	29.5	276.23
<i>P. vagneri</i>	80	7.2	646.23	15.4	39.5	276.23



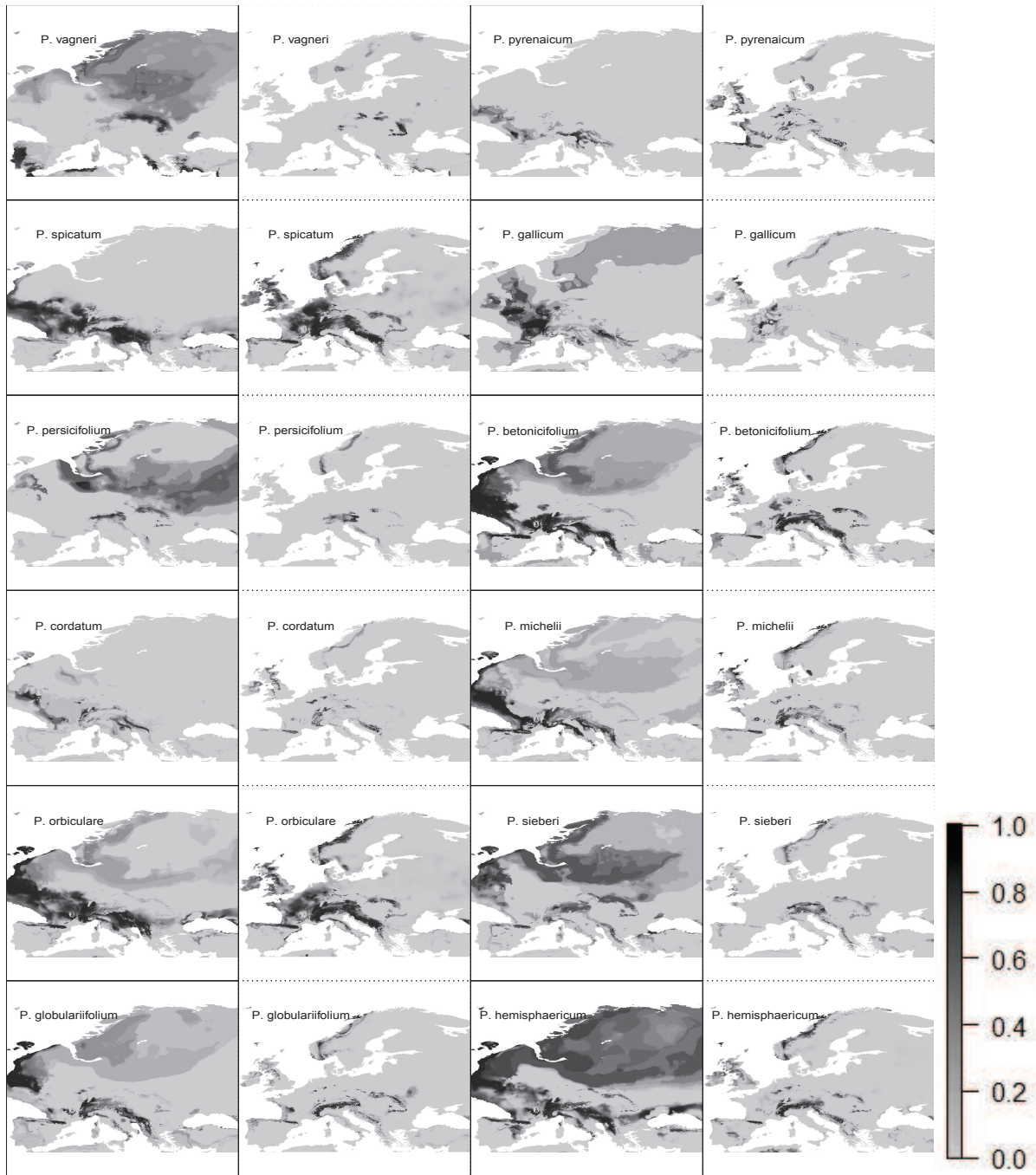


Figure 7: Projected habitat suitability obtained by the climatic niche modelling for all sister species for LGM climate (in boxes with black line) and current climate (in boxes with dotted line). The values range from 0 (climatic conditions expected not to be suitable for the species) to 1 (climatic conditions expected to be fully suitable for the species). It is obvious that the amount of suitable areas for species has decreased noticeably since the LGM.

### 3.3. Niche overlap

The one-tailed Wilcoxon rank sum test showed that sister species have a significantly higher niche overlap than non-sister species (Fig. 8). Amongst the sister species, *P. pyrenaicum* and *P. vagneri* stood out with no niche overlap at all. All other sister species had a niche overlap of around  $3.56 \pm 0.19$ . On average, the non-sister species had a lower niche overlap of  $0.21 \pm 0.16$ . 50 % of the range overlap was even lower at around 0.17.

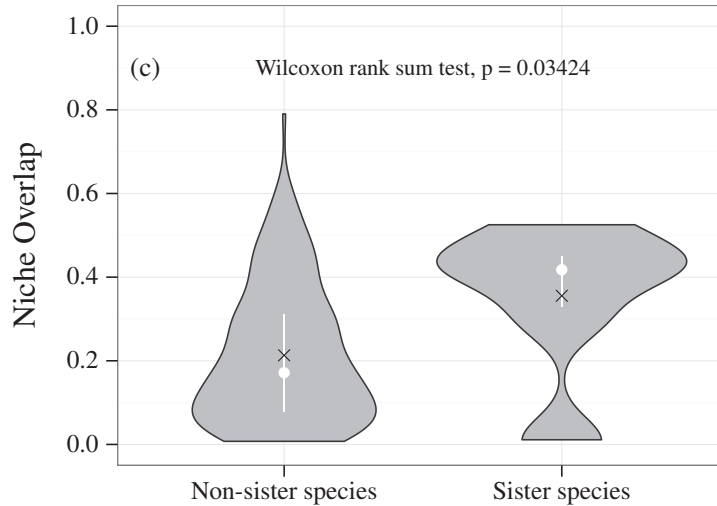


Figure 8: Violinplot of niche overlap for sister and non-sister species. The white line represents the 25% and 75% quantile, the white dot the median and the black cross the mean of the distribution. The p-Value results from a one-tailed Wilcoxon rank sum test.

### 3.4. Range overlap and range asymmetry

The maps in Fig. 9 depict the points of occurrence for all sister species. Contrary to our expectations, sister species did not have a significantly lower range overlap than non-sister species ( $p = 0.99$ ). We therefore repeated the one-tailed test in the opposite direction and the range overlap of sister species turned out to be higher than for non-sister species ( $p = 0.02$ ) (Fig. 10(a)). The range of *P. gallicum* lies completely within the range of *P. spicatum* (range overlap = 1). Likewise, the ranges of *P. cordatum* and *P. michelii* are almost identical (range overlap  $\approx 1$ ). *P. vagneri* and *P. pyrenaicum* on the contrary share no common territory at all (range overlap = 0). Half of the non-sister species pairs have no range overlap at all.

For simplicity, range asymmetry was plotted on log-scale because of the high values obtained. Sister and non-sister species differed significantly ( $p = 0.04$ ). Because *P. orbiculare* and *P. sieberi* showed a very high range asymmetry (63.55) compared to the other sister species (mean = 2.95), the Wilcoxon rank sum test was repeated without this value. No signi-

ficant difference between sister and non-sister species could be found anymore ( $p = 0.1348$ ) (Fig. 10(b)). Compared to the range asymmetry of *P. sieberi* and *P. orbiculare* the range asymmetry of *P. vagneri* and *P. pyrenaicum* was very low with  $\approx 0.71$ .

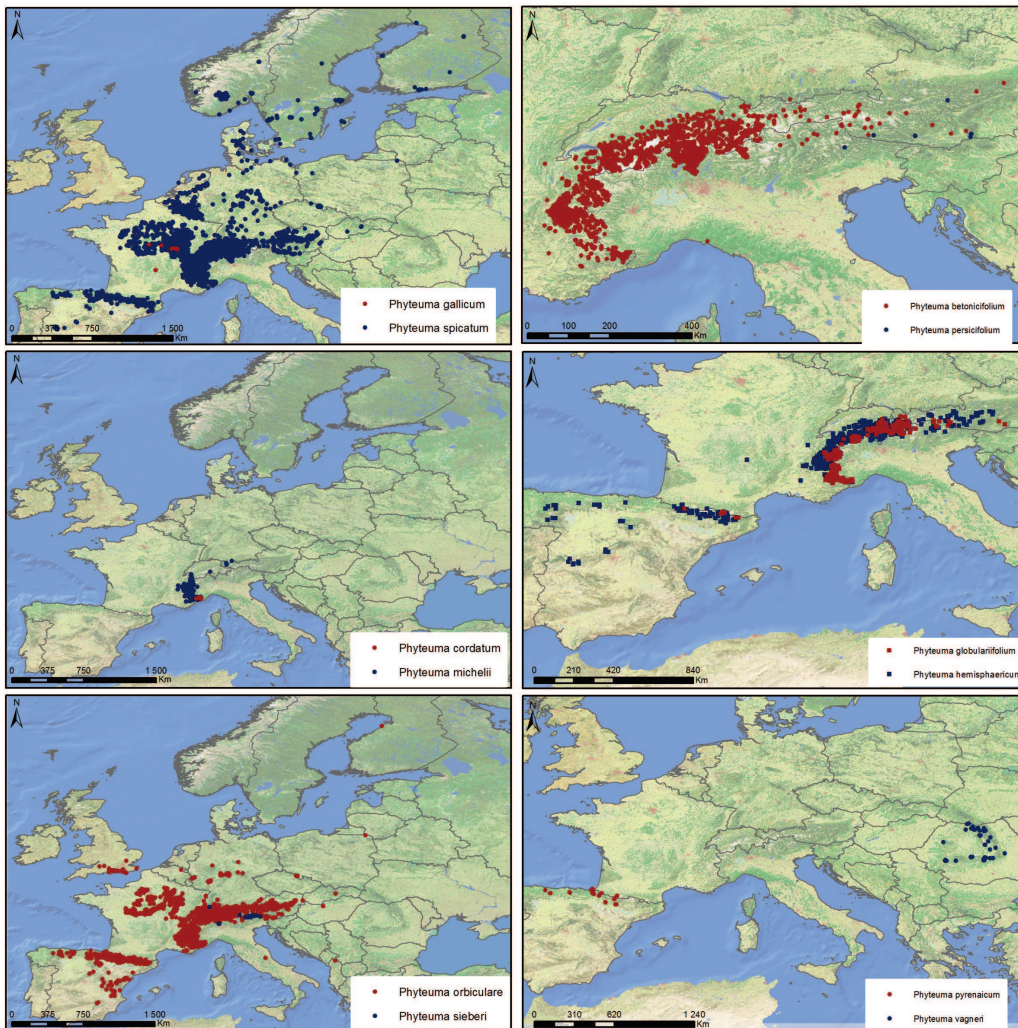


Figure 9: Maps of points of occurrence for sister species. The maps have been produced by Julien Renaud (LECA).

Generally, there was no significant relationship between the time since divergence of a species pair and range overlap for both sister species pairs ( $p = 0.99$ ) and non-sister species pairs ( $p \approx 0.47$ ). As the high range overlap of 1 between *P. spicatum* and *P. gallicum* appeared to be an outlier, it was removed from the regression to see whether the relationship between time since divergence and range overlap would change. Indeed, range overlap increases significantly with time since divergence of sister species without this species pair ( $p = 0.01$ ).

The one-tailed Wilcoxon signed rank test revealed that the area of distribution of *Phyteuma* could have been significantly higher in the past ( $p < 0.05$ ) (Fig. 12). As the projected ranges

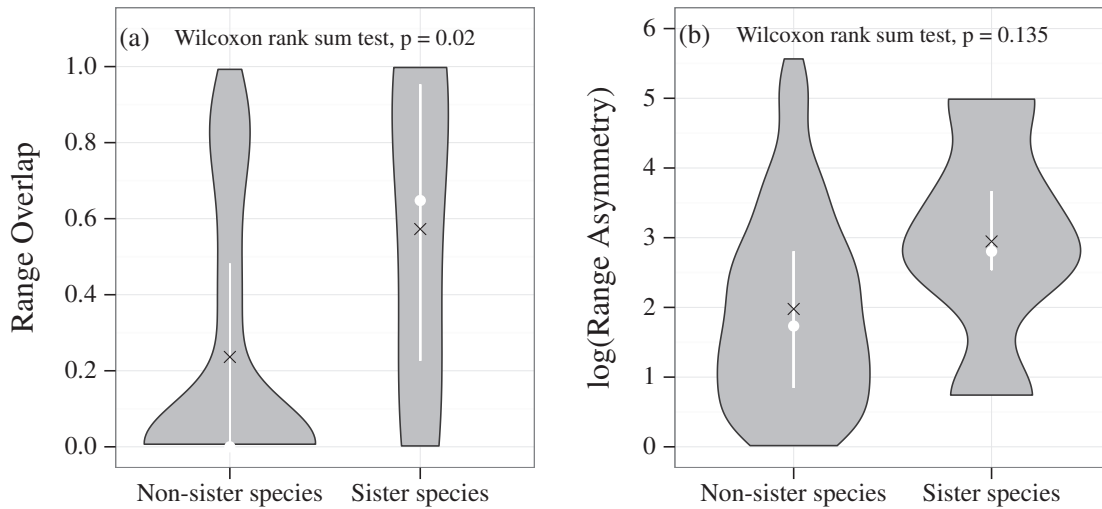


Figure 10: Violinplots of (a) range overlap and (b) range asymmetry for sister and non-sister species. For range overlap a one-tailed test was conducted, for range asymmetry a two-tailed test. For the range asymmetry violinplot and the Wilcoxon rank sum test presented in this plot the outlier of the sister species *P. orbiculare* and *P. sieberi* (range asymmetry = 63.55) was excluded so that only 5 sister pairs remained. The white line represents the 25% and 75% quantile, the white dot the median and the black cross the mean of the distribution.

were withdrawn from the habitat suitability projections from the CNMs, the high over-prediction of the CNMs needs to be taken into account for the interpretation of this result.

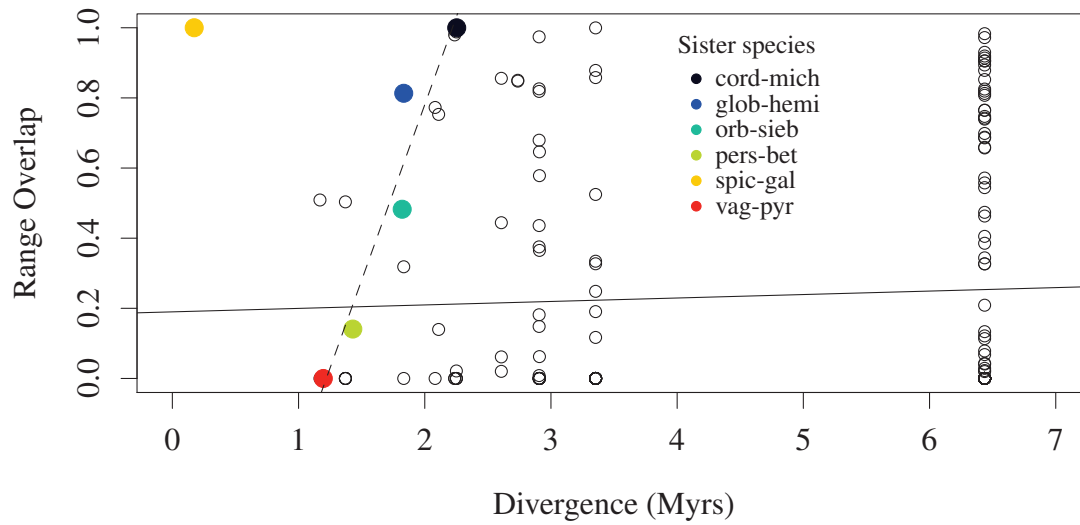


Figure 11: Range overlap as a function of time since divergence. Sister species have been emphasized by colouring. All other species pairs are represented by open circles. The linear regression resulted in a non-significant relationship for non-sister species ( $p \approx 0.47$ )(solid line) and a significant relationship for sister species once the species pair *P. spicatum* and *P. gallicum* was removed ( $p = 0.01$ ) (dashed line).

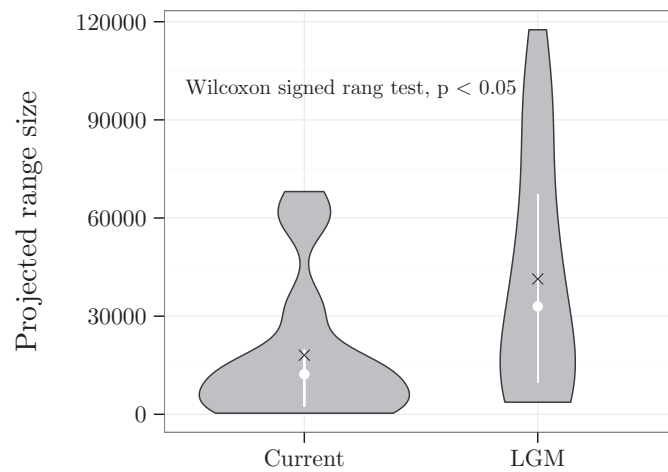


Figure 12: Violinplot of projected range size for current and LGM climate. The white line represents the 25% and 75% quantile, the white dot the median and the black cross the mean of the distribution. The p-value results from a one-tailed Wilcoxon signed rank test.



## 4. Discussion

### 4.1. Main findings and comparison to other studies

The results obtained largely support the hypothesis that *Phyteuma* mainly evolved due to geographic barriers that emerged in the course of environmental changes. The clear decrease of range size that can be expected to have occurred for most species since the Last Glacial Maximum and the estimated climatic niches obtained by the ecological niche modelling confirm the hypothesis that *Phyteuma* is a plant genus especially adapted to alpine climatic conditions. During times of glaciation it might have been widely distributed throughout Europe but the current warm period of the quaternary Ice Age forces it to occupy predominantly alpine regions. Schulz (1904) already commented on this when he tried to reconstruct the history of *P. serratum*, a species that occurs exclusively in Corsica. The only explanation that Schulz could find for the separation of *P. serratum* from the other species is that ancestors of *P. serratum* reached Corsica over a land bridge after the change to a colder climate had suppressed subtropical plants and enabled the alpine ancestors of *P. serratum* to descend into the lowlands. The changing climates of the past forced *Phyteuma* to move along with its preferred alpine habitat and led to hugely fragmented habitats during warmer periods. The high degree of niche overlap within sister species compared to non-sister species indicates that the species of *Phyteuma* did not develop any new traits that could have changed their climatic niches but rather moved along with the optimal climate they were accustomed to (niche conservatism). Additionally, *Phyteuma* apparently does not have any obvious phenomenological traits that hint at ecological differentiation as the main driver of diversification. All species of *Phyteuma* have a very similar morphology and are of the same life form. Boucher et al. (2011) analysed the evolutionary history of the plant genus *Androsace* and discovered that the change to the cushion life form had significantly spurred diversification in this genus. It can be expected that in the course of climatic changes populations of the ancestors of *Phyteuma* got separated and new species developed in vicariance due to the interrupted gene flow and reproductive isolation. Peterson et al. (1999) obtained similar results in their study of birds, mammals and butterflies in the Isthmus of Tehuantepec. They observed that younger sister species shared more similar niches than older ones and based on these results concluded that for the species they observed, vicariance and following ecological evolutionary changes were the driving forces of diversification. The missing pattern in range asymmetry among sister and non-sister species suggests that populations got split randomly during periods of unsuitable climatic conditions.

The only result that contrasts our expectation of *Phyteuma* to have evolved in vicariance is the higher range overlap of sister species compared to non-sister species. An explanation can be found in Aguilée et al. (2012) who showed that secondary contacts after allopatric speciation, where species rejoin each other again when more favourable climates return, can erase the expected low range overlap. They favour the evolution of reproductive isolation (assortative mating), which reinforces incipient speciation. Barraclough and Vogler (2000) observed similar patterns in their study on examples of bird, fish and insect groups. They suspect that only a high number of real sympatric speciation events could be distinguished from patterns

of secondary contact. The constantly changing climatic conditions (e.g. glacial fluctuations) that *Phyteuma* was exposed to during its evolutionary history could definitely have caused secondary contacts. If secondary contacts happened after allopatric speciation events, it would be expected that range overlap increases with the time since divergence. For sister species our results show a clear positive relationship between time since divergence and range overlap which could be proof that secondary contacts lead to high range overlap of formerly allopatric species (Fig. 11). This result needs to be treated with caution, though, as one sister species pair was removed from the regression so that only five remained and statistical power therefore was even more reduced. Furthermore, *P. pyrenaicum* and *P. vagneri* appear to be a special sister pair, too, so that they would eventually as well have to be excluded from the regression which would further reduce the statistical significance.

Some species pairs were highlighted in our analysis as proof that it can be too simplistic to try to classify one speciation mode for a whole clade.

*P. vagneri* and *P. pyrenaicum* seem to have undergone a long distance dispersal event. They inhabit ranges that lie about 2,500 km apart and thus do not have any range overlap. Their low niche overlap could therefore be a result of ecological evolution after geographical separation. The fact that their phenology differs greatly supports this assumption (Fig. 2(c),(d)). It could be interesting to investigate these two species closer to find out about the causes of this migration as the seeds of *Phyteuma* are very small ( $\approx 1\text{mm}$ ) and naturally only show short dispersal ranges of max. 10m. (Schulz, 1904; Maier et al., 1999). Nathan et al. (2008) pointed out that animals such as birds or extreme weather events can spur long distance dispersal.

Another interesting species pair within *Phyteuma* are *P. spicatum* and *P. gallicum* (Fig. 2(a), (b)). Within a genus that appears to have been mainly imprinted by allopatric speciation, this species pair could be an example for “budding speciation”. The large range of *P. spicatum* completely includes the range of *P. gallicum* and this, in combination with the high range asymmetry between the two ranges and the recent divergence time, are just the conditions Anacker and Strauss (2014) predict for the signature of “budding speciation”. In their study they found the same tendencies for their study groups in the California Floristic Province (CFP). Barraclough and Vogler (2000) also found two species within their study groups which clearly stood out with high range overlaps compared to the other species which have been identified as mainly allopatric.

## 4.2. Limitations

Despite some insights into the evolutionary history of the plant genus *Phyteuma*, the reliability and meaningfulness of these results remain restricted. This is mainly due to limitations of the methods available in this area of research and the lack of sufficient data about the past.

The caveats of ecological niche modelling and its application in the context of phylogeography have been widely discussed in recent literature (Dormann, 2006; Austin, 2007; Nogués-Bravo, 2009; Alvarado-Serrano and Knowles, 2014). There are several reasons why the models of climatic niches formulated for this study should be considered carefully.

The models were calibrated at a coarse resolution of 2.5 arc-minutes. This scale was chosen because of the large distribution area of *Phyteuma* and availability of data. Plants are immensely influenced by their microhabitat (e.g. orientation, availability of light, small scale changes of soil type)(Anacker and Strauss, 2014). This is especially true in mountainous regions where landscape and habitats tend to change rapidly within short distances (Scherrer and Körner, 2010). The smoothed climatic layers provided by Worldclim cannot represent these small scale changes and might not be able to reproduce the climate actually encountered by the species. For the modelling of climatic niches of *Phyteuma* this is particularly critical as almost all of its species are predominantly alpine. The model could be highly improved if data were sufficiently available on a higher resolution.

The high quality of our models confirmed by their TSS scores shows that the models did well in the evaluation part (prediction of actual points of occurrences) but they tended to over-predict the area of occurrence for all species. The inaccuracy of our models might result from an insufficient number of environmental variables used for calibration (Nogués-Bravo, 2009). Even though a PCA was conducted to determine the climatic variables which appeared to control the distribution of *Phyteuma*, climatic variables which are biologically limiting for *Phyteuma* might have been erroneously excluded from the calibration. Furthermore, other ecological factors which might restrict the occurrence of *Phyteuma* have been completely left out in this study but could probably be fundamental to explain its distribution (Wiens and Graham, 2005). It could for example make sense to gather detailed information about soil preferences and include them into model calibration as species of *Phyteuma* differently prefer calciferous and silicious soils (Schulz, 1904). Furthermore, ecological niche modelling so far only includes the realized niche of a species. Biological interactions, which have so far been rarely incorporated into ENMs, might be an import factor.

The fact that *Phyteuma* is usually only adapted to short distance dispersal has not yet been taken into account for the calibration of the CNM. Except for some individuals, most of the species have not been able to reach the highly suitable areas in Scandinavia or Great Britain predicted by the CNM. Including dispersal capabilities of *Phyteuma* in the CNM could further decrease over-prediction (Peterson, 2011).

Further, the sampling of species affects the accuracy, too (Wilfried Thuiller, pers. comm.). It is likely that our databases represent only a portion of the actual species occurrences. Especially the GBIF database, where institutions and hobby botanists feed in their observations of individuals of a plant species, does not offer an extensive sampling of areas where species of *Phyteuma* are expected to occur. Therefore, it can be assumed that our dataset is incomplete and might lack occurrence points in the areas that have been identified as "highly suitable" by the CNM.

For simplicity, only one model was used to represent the climatic niche in this study. As Araújo and New (2007) proposed, it might make sense to repeat the ecological niche modelling with as many models as possible to obtain one robust ensemble model out of all of them.

There are also some limitations to the credibility of our projections on past climates. Reconstructions of the past climate are flawed with uncertainties themselves and might not be able to give us an accurate picture about how climatic variables were correlated for the species existing back then. For example the model predicted that species would occur in areas which



are known to have been covered by glaciers. Even if some populations might have managed to survive in these areas on so called “nunataks”, they could not have been as widespread on ice as predicted by the model. Species should therefore be restricted to occur in these areas by the incorporation of a layer with information about the extent of glaciers during the LGM.

Another caveat of this study is the size of the phylogeny that was used. Concerning statistical power it is difficult to make extensive conclusions from 20 species with occurrence data and only 6 sister species. Extinctions that could not be included in the phylogeny due to missing fossil records might also have left a noise on the patterns we are observing today.

## 5. Conclusions and future perspectives

This study lay the foundation for further investigation of the evolutionary history of the plant genus *Phyteuma*. Climatic changes towards warmer temperatures and conservatism of climatic niches have been identified as the causes of separation of populations of the predominantly alpine genus and as the trigger of allopatric speciation. *Phyteuma* would therefore be a genus potentially endangered by the current global warming as its opportunities to migrate to colder climates become reduced. Landscape changes as proposed in Aguilée et al. (2012), appear to have lead to secondary contact during colder climatic periods and probably to further diversification. But as some species seem to have undergone different ways of speciation, it requires further research to finally reveal the evolutionary history of *Phyteuma* and be able to make predictions about its future.

As explained in the discussion part, one aspect is to further improve the climatic model to be able to make more precise and reliable conclusions.

A more detailed examination of genetic similarity of different populations, as done in Kozak and Wiens (2006) could help to further clarify the past. First reconstructions of chromosome number that were conducted by Cristina Roquet and Schneeweiß et al. (2013) indicate that ploidy shifts happened could have happened each time populations got separated by a geographical barrier and possibly reinforced the divergence by creating reproductive isolation (Wood et al., 2009).

Furthermore, it could be interesting to investigate the biogeographic history of *Phyteuma* to find out more about its origins and possible migration routes. Two powerful methods have been developed recently by Landis et al. (2013) and Matzke (2013): bayArea and BioGeoBEARS. bayArea is a Bayesian framework that enables the user to incorporate a large number of distributional areas to infer ancestral species ranges. BioGeoBEARS includes a large number of different biogeographical models and enables founder effects.

# References

## Articles

**Aguilée, R., Claessen, D., Lambert, A. (2012):** Adaptive radiation driven by the interplay of eco-evolutionary landscape dynamics, *Evolution*, 67 (5), 1291-1306

**Alvarado-Serrano, D., Knowles, L.L. (2014):** Ecological niche models in phylogeographic studies: applications, advances and precautions, *Molecular Ecology Resources*, 14, 233-248

**Anacker B.L., Strauss S.Y. (2014):** The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society B*, 281: 20132980

**Araújo, M.B., New, M. (2007):** Ensemble forecasting of species distributions, *Trends in Ecology and Evolution*, 22 (1), 42-47

**Austin, M. (2002):** Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, *Ecological Modelling*, 157, 101-118

**Austin, M. (2007):** Species distribution models and ecological theory: A critical assessment and some possible new approaches, *Ecological Modelling*, 200, 1-19

**Barracough, T.G., Vogler, A.P. (2000):** Detecting the Geographical Pattern of Speciation from Species-Level Phylogenies, *The American Naturalist*, 155 (4), 419-434

**Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., Courchamp, F. (2012):** Impacts of climate change on the future of biodiversity, *Ecology Letters*, 15, 365-377

**Bird, C.E., Fernandez-Silva, I., Skillings, D.J., Toonen, R.J. (2012):** Sympatric Speciation in the Post "Modern Synthesis" Era of Evolutionary Biology, *Evolutionary Biology*, 39 (2), 158-180

**Boucher, F.C., Thuiller, W., Roquet, C., Douzet, R., Aubert, S., Alvarez, N., Lavergne, S. (2011):** Reconstructing the origins of high-alpine niches and cushion life form in the genus *Androsace* S.L. (Primulaceae), *Evolution*, 66 (4), 1255- 1268

**Diniz Filho, J.A.F., Ferro, V.G., Santos, T., Nabout, J.C., Dobrovolski, R., De Marco Jr., P. (2010):** The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae), *Revista Brasileira de Entomologia*, 54 (3), 339-349

**Dormann, C. (2006):** Promising the future? Global change projections of species distributions, *Basic and Applied Ecology*, 8, 387-397

**Evans, M.E.K., Smith, S.A., Flynn, R.S., Donoghue, M.J. (2009):** Climate, Niche Evolution, and Diversification of the “Bird-Cage” Evening Primroses (*Oenothera*, Sections *Anogra* and *Kleinia*), *The American Naturalist*, 173 (2), 225-240

**Fitzpatrick, B.M., Turelli, M. (2006):** The geography of mammalian speciation: Mixed signals from phylogenies and range maps, *Evolution*, 60 (3), 601-615

**Gavrilets, S., Li, H., Vose, M.D. (2000):** Patterns of parapatric speciation, *Evolution*, 54 (4), 1126-1134

**Grossenbacher, D.L., Veloz, S.D., Sexton, J.P. (2014):** Niche and range size patterns suggest that speciation begins in small ecologically diverged populations in North American Monkeyflowers (*Mimulus* spp.), *Evolution*, 68 (5), 1270-1280

**Guisan, A., Zimmermann, N.E. (2000):** Predictive habitat distribution models in ecology, 135, 147-186

**Guisan, A., Thuiller, W. (2005):** Predicting species distribution: offering more than simple habitat models, *Ecology Letters*, 8, 993-1009

**Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L. (2006):** The effect of sample size and species characteristics on performance of different species distribution modeling methods, *Ecography*, 29, 773-785

**Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A. (2005):** Very high resolution interpolated climate surfaces for global land areas, *International Journal of Climatology*, 25, 1965- 1978

**Hoffmann, A.A., Sgrò, C.M. (2011):** Climate change and evolutionary adaptation, *nature*, 470 (7335), 479-485

**Hutchinson, G.E. (1957):** Concluding remarks, *Cold Spring Harbor Symposia on Quantitative Biology*, 22 (2), 415–427

**Kozak, K.H., Wiens, J.J. (2006):** Does niche conservatism promote speciation? A case study in North American salamanders, *Evolution*, 60 (12), 2604-2621

**Kozak, K.H., Graham, C.H., Wiens, J.J. (2008):** Integrating GIS-based environmental data into evolutionary biology, *Trends in Ecology and Evolution*, 23 (3), 141-148

**Landis, M.J., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P. (2013):** Bayesian Analysis of Biogeography when the Number of Areas is Large, *Systematic Biology*, 62 (6), 789-804

**Maier**, A., Emig, W., Leins, P. (1999): Dispersal patterns of some *Phyteuma* species (Campanulaceae), *Plant Biology*, 1, 408-417

**Marmion**, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W. (2009): Evaluation of consensus methods in predictive species distribution modelling, *Diversity and Distributions*, 15, 59-69

**Matzke**, N.J. (2013): Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing, *Frontiers of Biogeography*, 5 (4), 242-248

**Nathan**, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A., Tsoar, A. (2008): Mechanisms of long-distance dispersal, *Trends in Ecology and Evolution*, 23 (11), 538-647

**Nogués-Bravo**, D. (2009): Predicting the past distribution of species climatic niches, *Global Ecology and Biogeography*, 18, 521-531

**Olesen**, J.M., Alarcón, M., Ehlers, B.K., Aldasoro, J.J., Roquet, C. (2012): Pollination, biogeography and phylogeny of oceanic island bellflowers (Campanulaceae), *Perspectives in Plant Ecology, Evolution and Systematics*, 14, 169-182

**Paiva**, M.R., Santos, H., Kerdelhué, C., Mateus, E.P., Branco, M.R. (2011): Can climate change drive speciation?, *Proceedings of the Global Conference of Global Warming 2011*, 11-14 July 2011, Lisbon, Portugal

**Peterson**, A.T. (1999): Conservatism of Ecological Niches in Evolutionary Time, *Science*, 285 (5431), 1265-1256

**Peterson**, A. T. (2011): Ecological niche conservatism: a time-structured review of evidence, *Journal of Biogeography*, 38, 817-827

**Quézel**, P. (1954): Contribution à la flore de l'Afrique du Nord: 3. Contribution à la flore du J. Bou-Naceur (Moyen Atlas oriental), *Bulletin de la Société d'histoire naturelle d'Afrique du Nord*, 44, 255-263

**R Core Team** (2014): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [www.R-project.org/](http://www.R-project.org/)

**Rundle**, H.D., Nosil, P. (2005): Ecological speciation, *Ecology Letters*, 8, 336-352

**Salichos**, L., Stamatakis, A., Rokas, A. (2014): Novel information theory-based measures for quantifying incongruence among phylogenetic trees, *Molecular Biology and Evolution*, 31, 1261-1271

**Sanderson, M.J.** (2003): r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock, *Bioinformatics*, 19, 301-302

**Scherrer, D., Körner, C.** (2010): Infra-red thermometry of alpine landscapes challenges climatic warming projections, 16, 2602-2613

**Schneeweiss, G.M., Paschwöll, C., Tribsch, A., Schönswetter, P., Barfuss, M.H.J., Esfeld, K., Weiss-Schneeweiss, H., Thiv, M.** (2013): Molecular phylogenetic analyses identify Alpine differentiation and dysploid chromosome number changes as major forces for the evolution of the European endemic *Phyteuma* (Campanulaceae), *Molecular Phylogenetics and Evolution*, 69, 634- 652

**Schnitzler, J., Graham, C.H., Dormann, C.F., Schiffers, K., Linder, H.P.** (2012): Climatic niche evolution and species diversification in the Cape flora South Africa, *Journal of Biogeography*, 39, 2201-2211

**Schoener, T.W.** (1968): *Anolis* lizards of Bimini: resource partitioning in a complex fauna, *Ecology*, 49, 704-726

**Stamatakis, A.** (2014): RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*, 30, 1312-1313

**Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Skyes, M.T., Zimmermann, N.E.** (2008): Predicting global change impacts on plant species' distributions: Future challenges, *Perspectives in Plant Ecology, Evolution and Systematics*, 9, 137-152

**Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B.** (2009): BIOMOD – a platform for ensemble forecasting of species distributions, *Ecography*, 32, 369-373

**Turelli, M., Barton, N.H., Coyne, J.A.** (2001): Theory and speciation, *TRENDS in Ecology & Evolution*, 16 (7), 330- 343

**Warren, DL., Glor, RE., Turelli, M.** (2008): Environmental Niche Equivalency versus conservatism: Quantative approaches to niche evolution, *Evolution*, 62, 2868-2883

**Wiens, J.J., Graham, C.H.** (2005): Niche Conservatism: Integrating Evolution, Ecology and Conservation Biology, *Annual Review of Ecology, Evolution and Systematics*, 36, 519-539

**Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H.** (2009): The frequency of polyploid speciation in vascular plants, *Proceedings of the National Academy of Sciences of the USA*, 106 (13), 13875–13879

**Zimmermann, N.E.** (2009): update: More uncertainty with BIOMOD, *Frontiers of Biogeography*, 1 (2), 38-39

## **Books**

**Dormann, C.F.** (2013): *Parametrische Statistik*, Springer, Berlin/Heidelberg, 350 p.

**Landolt, E.** (2010): *Flora indicativa*, Haupt, Bern, 378 p.

**Lauber, K., Wagner, G.** (2007): *Flora Helvetica*, 4.Auflage, Haupt, Bern, 1629 p.

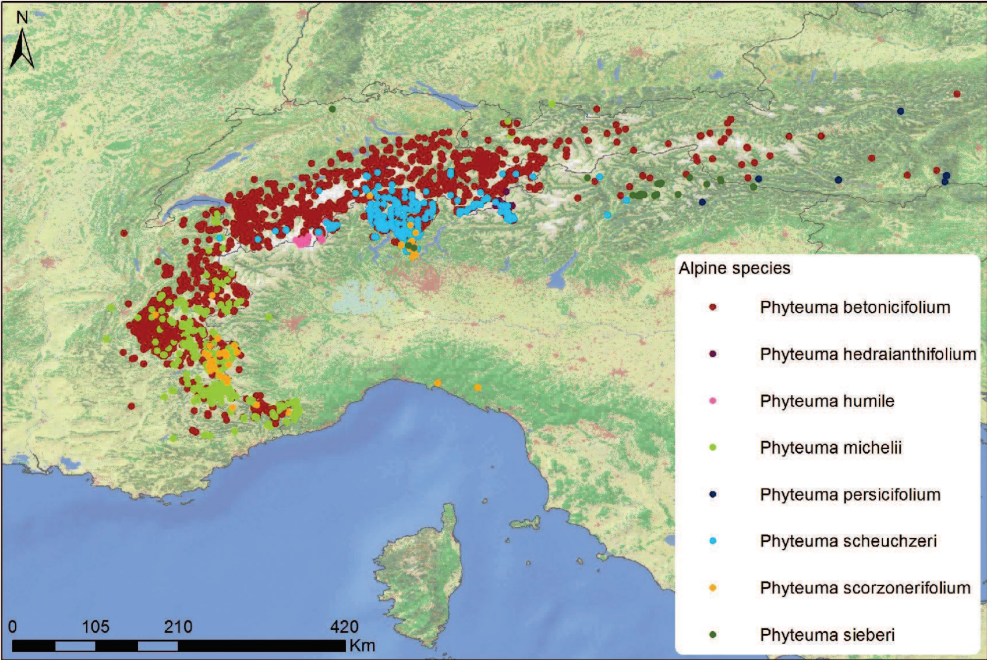
**Schulz, R.** (1904): *Monographie der Gattung Phyteuma*, Kessinger Legacy Reprints, Whitefish, 204 p.

**Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M.** (2009): *Mixed Effects Models and Extensions in Ecology with R*, Springer, New York, 574 p.

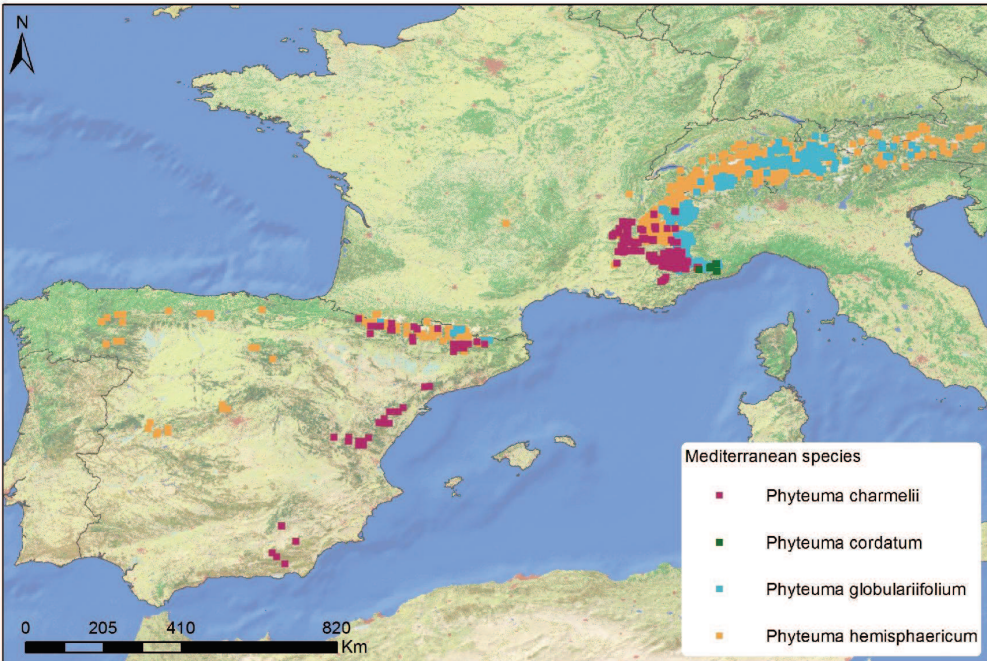
# Appendices



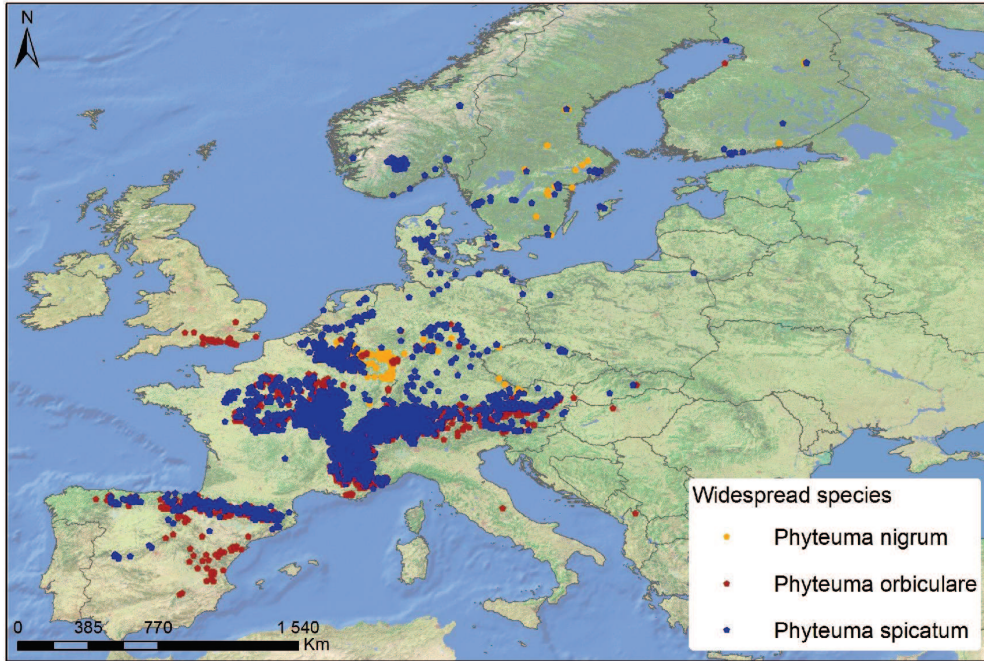
# A. Information occurrences *Phyteuma*



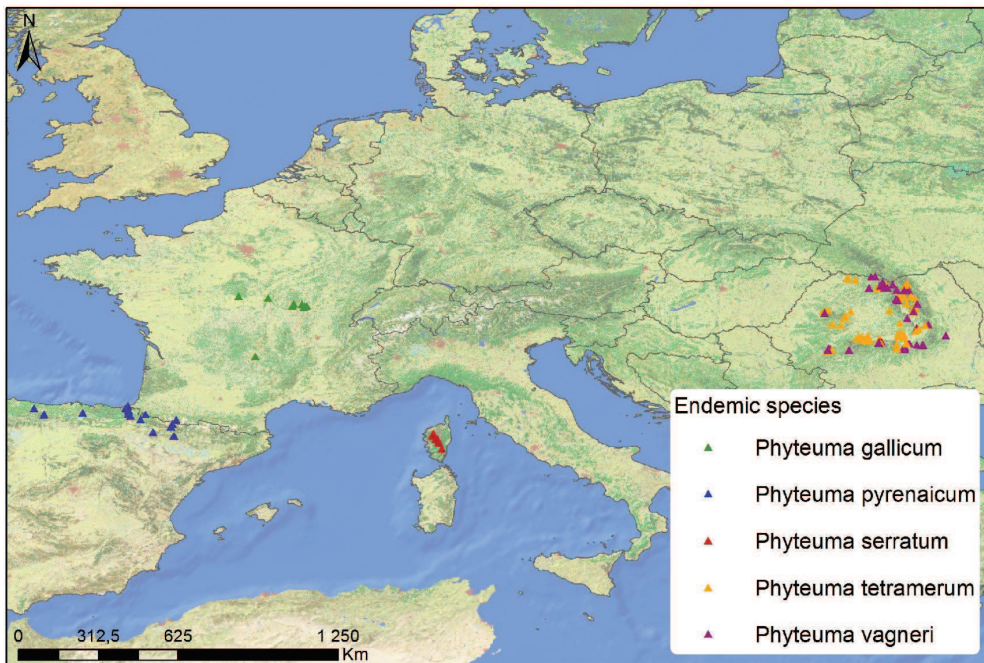
(a)



(b)



(c)



(d)

Figure A.1: Maps of points of occurrence of *Phyteuma* obtained from several databases. The maps have been produced by Julien Renaud (LECA).

- (a) Distribution of Alpine species,
- (b) distribution of Mediterranean species,
- (c) distribution of widespread species and
- (d) distribution of endemic species



## B. Phylogeny

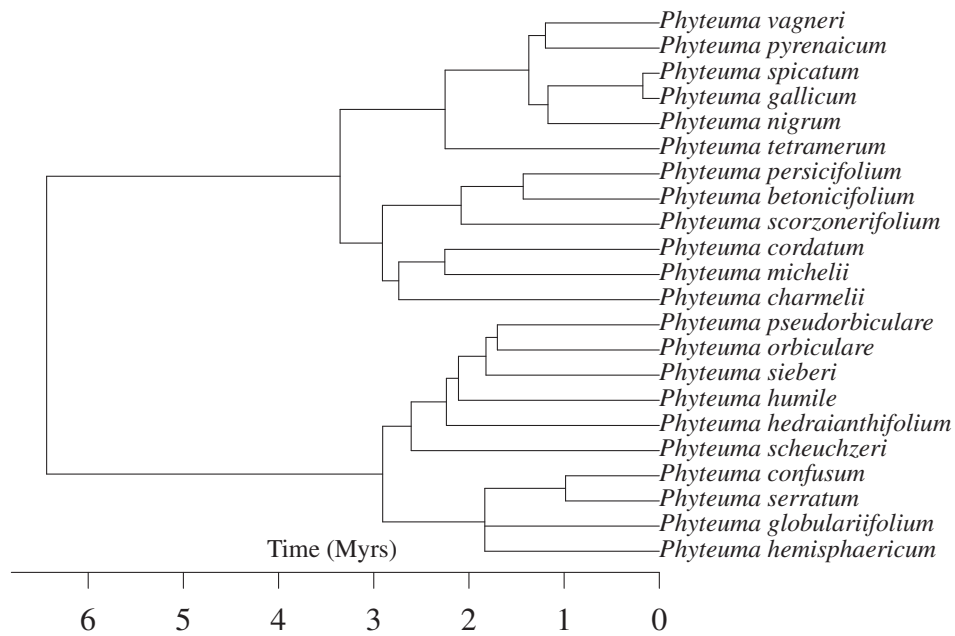


Figure B.1: Best maximum likelihood tree obtained by the phylogenetic inference. The branches are dated and the scalebar represents time in Million years

### C. Climatic niche modeling

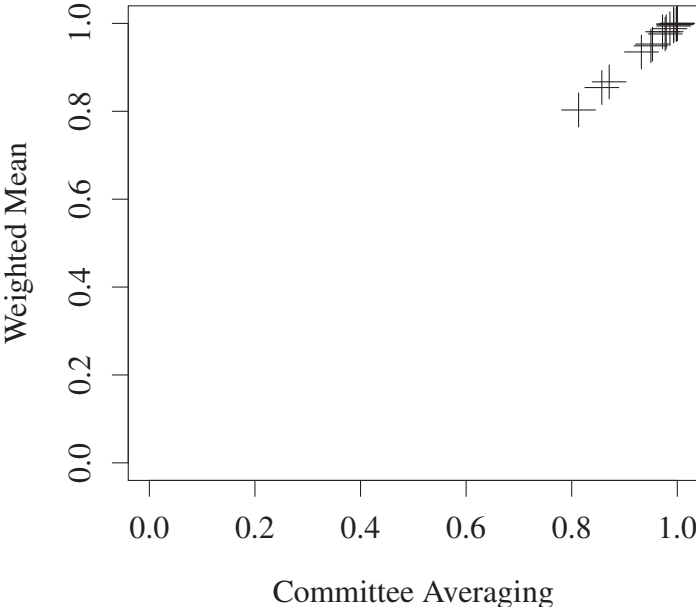


Figure C.1: Correlation of TSS scores of ensemble models produced by weighted mean probabilities and committee averaging