Evolution and Phylogeography of the North American genus *Boechera* (Brassicaceae) and the Evolution of Apomixis

Dissertation

Christiane Kiefer
born in Stuttgart

2008

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Diplom-Biologin Christiane Kiefer
born in Stuttgart/Germany
Oral-examination:: .....

Evolution and Phylogeography of the

North American genus *Boechera* (Brassicaceae)

and the Evolution of Apomixis

Referees:

Prof. Dr. Marcus Koch

Prof. Dr. Thomas Rausch

My PhD thesis is dedicated to my parents

Erich Kiefer 09.12.1939 – 21.07.2007

“Nichts ist so schlecht, dass es nicht für etwas anderes wieder gut ist”
(Nihil adeo est malum, non sit aliquid boni)
He taught me to see the bright side – no matter what.

Karin Kiefer geb. Nagel

She saw the interest for biology in me when I was a little child.
She constantly supported my hunger for knowledge on nature.

My life is dedicated to Science

Der unermesslich reichen, stets sich erneuernden Natur gegenüber wird der Mensch, soweit er auch in der wissenschaftlichen Erkenntnis fortgeschritten sein mag, immer das sich wundernde Kind bleiben und muss sich stets auf neue Überraschungen gefasst machen.

(*No matter how far humans progress in their scientific understanding, they will always remain marvelling children in the face of the infinite wealth and continual change within nature, and will always have to be ready for new surprises.*)

Max Planck (1858 - 1947)

Table of contents

Supplementary data from the chapters, a distribution map showing all accessions used in this thesis and an electronic copy of the thesis itself are given on the DVD in the back of the book.

0.1 Summary

Boechera is a North American genus of the Brassicaceae named after the Danish botanist Tyge Böcher who cytogenetically studied members of the genus in great detail. Until the end of the 20th century most species which belong today to *Boechera* were still included in the genus *Arabis*. Molecular studies in the late 1990s revealed that *Arabis* was polyphyletic and subsequently a majority of North American *Arabis* species was transferred into *Boechera*.
Boechera inhabits a wide range of habitats and reproduces sexually as well as apomictic (asexual reproduction via seeds).
The aims of the study presented inhere were (a) to reconstruct the phylogeographic history of the genus based on chloroplast DNA marker sequences and (b) to investigate the nrDNA ITS gene pool and reconstruct phylogenies based on nrDNA ITS and single copy gene introns.
In the course of the experiments it appeared that eastern North American *Boechera* constitute different evolutionary lineages than the species centred in western North America. Hence the split of eastern and western North American *Boechera* species became a third subject (c).
In our continental wide phylogeographic studies we detected a large amount of haplotype sharing indicating recurrent hybridisation on the one hand and non-differentiation of haplotypes since speciation on the other hand. We concluded that the chloroplast gene pool in *Boechera* pre-dates speciation in respect to the investigated markers. Unrelated from taxon identity we could show that the evolutionary lineages detected have a different phylogeographic history in terms of glacial refugia and recently recolonized areas.
The study based on nrDNA ITS and introns of two single copy genes enabled us to obtain deeper insights into *Boechera* phylogeny and ITS type distribution across taxa. We could show that species specific lineages exist although the relationship among them is poorly resolved. Comparing gene tree topologies this indicates rapid speciation which probably happened in the second half of the quaternary. Hybrids could be identified by the comparison of the different marker systems together with chloroplast DNA types from an earlier study.
The comparison of eastern and western North American *Boechera* based on DNA marker sequences showed that eastern North American *Boechera* represent two evolutionary lineages within the genus *Boechera* (cpDNA) or among the tribe Boechereae (nrDNA). We also included the Siberian taxa *Borodinia* and *Boechera falcata* in the analysis and could show that they were resolved within *Boechera*.
The final task in the thesis was to unravel the origin of two apomict specific chromosomes (Het and Del). BAC-FISH analyses revealed that Het is a homologue of *Boechera stricta* linkage group 1 and Del appears only in 15 chromosome apomicts and is a fragment of Het.
The thesis presented inhere offers deeper insights into the evolution of this complex genus and offers a good basis for ongoing and future research projects dealing with evolution and expression of apomixis as well as genome evolution in *Boechera*.

0.2 Zusammenfassung

Boechera ist eine nordamerikanische Gattung aus den Brassicaceae. Die Gattung wurde nach dem dänischen Botaniker Tyge Böcher benannt welcher sie in den 1950er Jahren cytogenetisch untersuchte. Bis zum Ende des 20. Jahrhunderts wurde *Boechera* zur Gattung *Arabis* gezählt, bis molekularbiologisch gezeigt werden konnte, dass *Arabis* polyphyletisch ist. Im Folgenden wurden daher Taxa welche bisher zu *Arabis* gerechnet wurden in die bereits 1971 beschriebene Gattung *Boechera* überführt. *Boechera* lebt in einer großen Anzahl verschiedener Habitate und kann sich auf sexuellem wie auch apomiktischen Wege (asexuelle Fortpflanzung über Samen) fortpflanzen.
Die Ziele der vorliegenden Studie waren (a) die phylogeographische Geschichte der Gattung basierend auf chloroplastidären Markersequenzen zu rekonstruieren und (b) den nrDNA ITS Genpool zu untersuchen und phylogenetische Untersuchungen basierend auf nrDNA ITS und Single Copy Gen Intronsequenzen zu rekonstruieren.
Im Verlauf der Experimente wurde deutlich, dass nordostamerikanische *Boechera* Arten in anderen evolutionären Linien zu finden sind als die Arten, welche ihren Verbreitungsschwerpunkt im westlichen Nordamerika haben. Daher ergab sich als dritter Projektteil (c) die Untersuchung der phylogenetischen Beziehungen der beiden Artengruppen.
In der kontinentweiten phylogeographischen Studie konnte vermehrt Haplotyp Sharing festgestellt werden welches entweder als Indikator fortwährender Hybridisierung oder nicht-Differenzierung de Haplotypen seit Artbildung (in Bezug auf den verwendeten Marker) interpretiert werden kann. Es folgt, dass die Differenzierung des Chloroplasten Genpools vor der Artbildung stattfand. Unabhängig von Taxonidentität konnte gezeigt werden, dass die detektierten evolutionären Linien eine unterschiedliche phylogeografische Vergangenheit in Bezug auf Refugialgebiete und nacheiszeitlich wieder besiedelte Gebiete haben.
Die auf nrDNA ITS und Single Copy Gen Intronen basierende Studie gestattete es einen tieferen Einblick in die Phylogenie und Verteilung von ITS Typen in einzelnen Taxa zu erhalten. Es konnte gezeigt werden, dass artspezifische Linien existieren obwohl ihre Beziehung zueinander nicht aufgelöst werden kann. Der Vergleich der Topologie unterschiedlicher Genbäume implizierte, dass die Differenzierung unterschiedlicher Linien über einen kurzen Zeitraum in der zweiten Hälfte des Quartärs erfolgte. Hybriden konnten durch den Vergleich der unterschiedlichen Markersysteme mit Chloroplasten-DNA Haplotypen aus dem ersten Projektteil identifiziert werden.
Der Vergleich NW- und NO –Amerikanischer *Boechera* Arten basierend auf DNA-Marker Sequenzen ergab, dass beide Artengruppen in unterschiedlichen evolutionären Linien zu finden sind. Es konnte ebenfalls gezeigt werden, dass die sibirischen Taxa *Borodinia* und *Boechera falcata* beide innerhalb der Gattung *Boechera* stehen.
Schlussendlich konnte mittels BAC-FISH Analysen die Herkunft zweier für Apomikten spezifischen Chromosomen geklärt werden (Het und Del). Het ist ein Homolog der *Boechera stricta* Kopplungsgruppe 1 und Del ist ein Bruchstück davon welches in Apomikten mit 15 Chromosomen vorkommt.
Insgesamt gibt die in diesem Rahmen vorgestellte Arbeit tiefere Einblicke in die Evolution dieser komplexen Gattung und bietet eine Basis für momentane und zukünftige Projekte zur Erforschung der Evolution und Expression von Apomixis sowie Genomevolution in der Gattung *Boechera*.

1. Introduction

1.1 Phylogeography

1.1.1 An Introduction to Phylogeography

The quaternary period beginning 1.8 million years ago has been marked by temperature oscillations caused by different inclinations of Earth towards the sun (Milankovic theory, Seibold and Seibold, 2005). During periods in which global temperatures decreased ice sheets extended from the poles into lower latitudes and in montaneous regions from higher to lower altitude. The coverage of landmasses by ice shields and the general drop in temperature had a major impact on fauna and flora. Animals and plants were forced to retreat into glacial refugia until rising temperatures allowed them to recolonize their former distribution ranges . The effect of those quaternary migrations is visible until today. They influenced along with habitat preferences the formation of present species distribution (Hewitt, 200, Hewitt, 2004).

Species distribution alone may be studied in biogeography. But how are geographical distribution and phylogeny related? This is precisely what phylogeographic studies try to answer. The term phylogeography itself was introduced in 1987 by John Avise (Avise, 1987). Since then a multitude of studies revealed migration patterns of plants and animals allover the globe (Avise, 1998). Phylogeographic inference makes use of the fact that quaternary migrations not only left their footprints in species distribution but also in the distribution of DNA sequence types. Typically plant studies employ non-coding pieces of chloroplast DNA as so called marker systems. Chloroplast DNA is a desirable molecule in angiosperms since in most cases it is uniparentaly (maternally) inherited (Birky, 1995) and the “migration of the seed” may be followed. The relatedness of DNA sequences calculated from mutations together with the abundance of different DNA sequence types in a region enable the inference of possible phylogeographic scenarios.

1.1.2 Ice Ages and Vegetation History in North America

During the quaternary North America was affected by two main glacials termed Illinoian and Wisconsin. The Illinoian includes two periods of glaciation and lasted from 300,000 to 130,000 years ago (Liesiecki et al. 2005). During the Illinoian the Laurentide ice shield covered about 85% of the area of Illinois and reached into Kansas during the maximum extent (Stiff and Hansel, 2004). The Sangamon or Eemian interglacial is temporarily equivalent to the Ipswichian Stage in Great Britain and the Riss-Würm interglacial in the European Alps. It began about 128,000-130,000 years ago. Northern hemisphere winters were slightly warmer and wetter than today. By 115,000 years ago the glacial era had returned (Lang, 1994). The last glacial period was the Wisconsin period which started about 100,000 years ago and

ended between 10,000 and 15,000 years ago. Glaciation reached its maximum extent by 18,000 years ago and affected mainly the northern but to some extent also the southern hemisphere. Canada and the northern United States were almost completely covered by the Laurentide ice-sheet while Alaska remained mostly ice-free (Adams and Faure, 1997). Local glaciations existed in the Rocky Mountains and the Cordilleran ice shield as well as in the Sierra Nevada in northern California (ice fields and ice caps, James et al. 2002).
The so-called Pinedale or Fraser glaciation was the last glaciation which affected the Rocky Mountains. It lasted from approximately 30,000 to 10,000 years ago (Brief geologic history, Rocky Mountain National Park). It mainly consisted of mountain glaciers merging into the Cordilleran ice shield (Ice Age Floods, From: U.S. National Park Service Website). The Cordilleran ice shield itself was responsible for phenomena like the glacial lake Missoula which cause the Missoula floodings after breakage of the ice dam closing the lake (about 40 times between 15,000 and 13,000 years ago) (Waitt, 1985). The Wisconsin glacial period which refers to the last extension of continental glaciers in the Laurentide ice shield had three glacial maxima called Tahoe, Tenaya and Tioga. The Tahoe reached its maximum by 70,000 years ago and the Tioga began about 30,000 years ago and ended 10,000 years ago. At the hight of glaciation the Bering Strait was covered by ice and allowed migration of humans and mammals. At the hight of the Wisconsin glaciation Canada, the Upper Midwest and New England as well as parts of Montana and Washington were covered in ice. In southwestern Saskatchewan and southeastern Alberta the Laurentide and Cordilleran ice shield met and formed the cypress hills. The Great Lakes resulted from pooled water from the melting glaciers (Damery, 2004).

With the quaternary climatic oscillations vegetation changed dramatically in North America as can be seen from pollen records. About 40,000 years ago (C^{14} time estimate) spruce and jack pine forests seemed to cover most of the eastern United States and for Tennessee and North Carolina mixed, temperate forests are reported. South of this in Texas southern pine forest with oak an hickory is reported. Eastern Canada was still covered by an ice-shield extending as far as the Great Lakes (Delcourt and Delcourt, 1981). 25,000 to 28,000 years ago, shortly before the last glacial maximum most of the eastern United States possibly had an open woodland vegetation. A mixed cool temperate forest belt seems to have existed across the southern Appalachian Mountains.
In the western Cordillera lake levels were higher but had not reached their maximum hight yet. In northern Arizona vegetation belts had already declined. In the north western Cordillera forest cover was less than today with more steppe and cold-tolerant species (Whitlock and Bartlein, 1997).

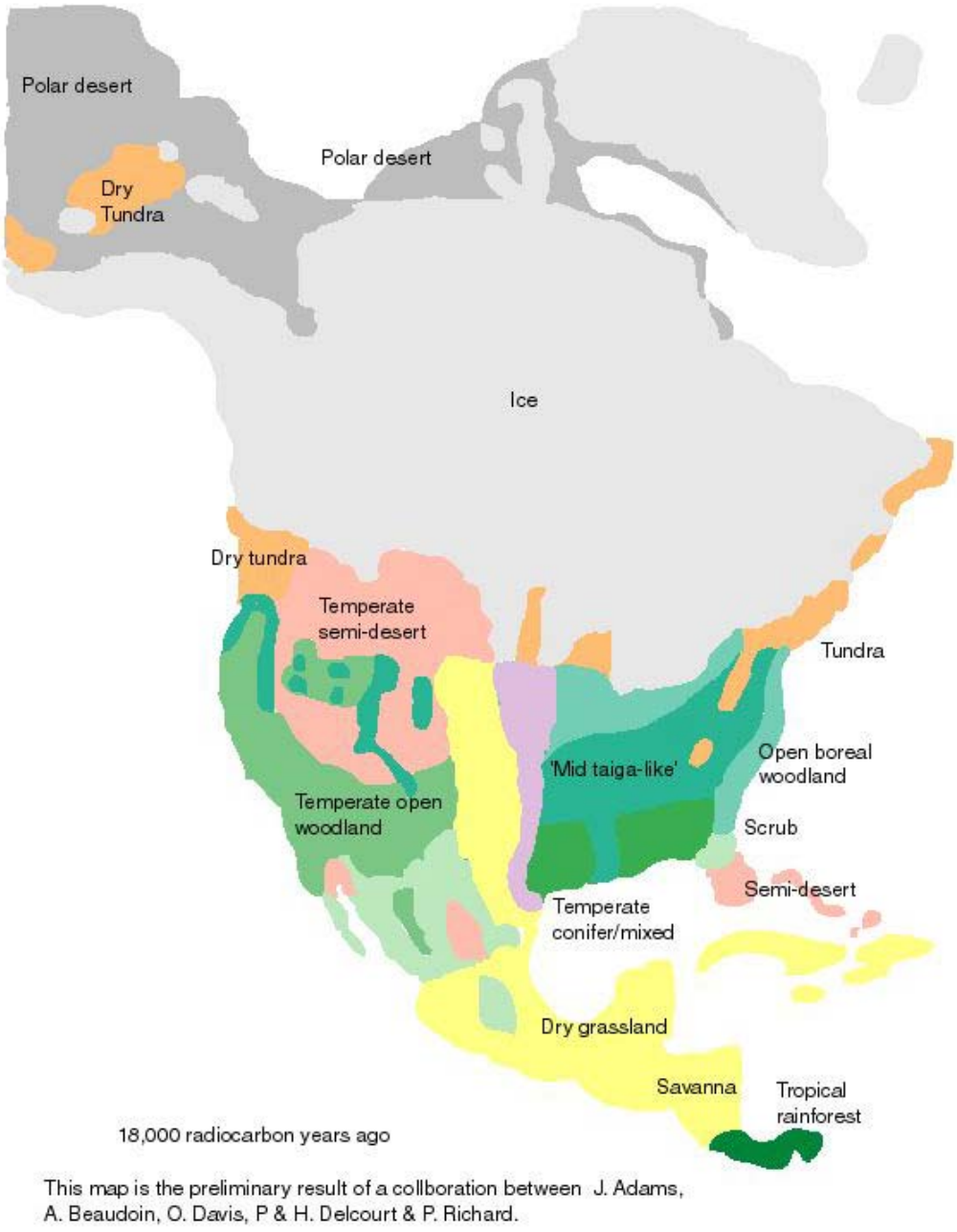

Fig. 1.1 Vegetation in North America during the last glacial maximum as given in Adams and Faure, 1997.

Vegetation types for the period during the last glacial maximum (18,000-15,000 C14 years ago; 21,000-17,000 calender years ago) are given in figure 1.1 according to Adams and Faure, 1997. Most of Canada (including the islands along the west coast) and the Northern United States were covered by an ice shield. Alaska remained ice free and was covered by polar desert and dry tundra in the central part (Ritchie & Cwynar 1982). In eastern North America open woodlands with pine and spruce replaced the present deciduous forests (Jackson et al. 1997). Open spruce woodland

Fig. 1.2 Present day potential vegetation in North America according to Adams and Faure, 1997.

extended further west into the prairie zone. The north-west of the United States was dominated by dry alpine tundra and polar desert (Whitlock and Bartlein, 1997) with scattered areas of cold-tolerant conifer woodland in the central Rocky Mountains. Further south in the Cordillera a mosaic of semi-desert scrub and sparse conifer woodland was found (Thompson et al. 1993). The south-west of North America was moister than today. Open conifer woodlands and scrub covered areas which are semi-desert today (for comparison of vegetation types compare figure 1.1 and figure 1.2).

Since 3000 years ago vegetation developed towards what is found today in North America.

1.1.3 Present Day Ecoregions of North America

Landscape may be subdivided into ecoregions which are characterized by a certain type of flora, fauna and environmental factors. As a result of several mapping projects the land masses of Earth have been subdivided into 867 ecoregions of which 116 are located in North America. Analyses in the present study were based on ecoregions as described in Ricketts et al. 1999. In the following details on the larger ecoregions which play a major role in the study and an outline on present vegetation is given. Representative pictures are given in figures 1.4 and 1.5.

Temperate Broadleaf and Mixed Forests

New England/Acadian Forest

A hilly to montaneous area in Eastern Canada with a mosaic of forest types and non-forest habitats. Northern hardwood and spruce forests predominate. Wide distribution of *Picea rubens* and *Pinus resinosa* distinguish this ecoregion from the adjacent Great Lakes ecoregion with predominantly deciduous forest.

Glaciers shaped the topography of this landscape with its characteristic mountains and plateaus. Soils are mostly poor and swamps and lakes are frequent. Summers are warm and moist, winters are cold and snowy.

Great Lakes Forests (East and West)

The climate in this ecoregion is affected by the Great Lakes resulting in longer growing season and effects on average temperature as well as timing and amount of precipitation. The region was entirely covered in glaciers and the ground is covered by a thick layer of deposits of glacial drift.

The Western Great Lakes region is characterized by mixed forests with *Populus tremuloides*, *Betula papyrifera*, *Pinus banksiana*, *Picea glauca*, *Picea mariana* and *Abies balsamea*. Wetlands are widespread in this region.

The Eastern Great Lakes forests lie between the boreal forests and the broadleaf deciduous forests and are therefore transitional. Parts of the forests are coniferous while others form a mosaic of pure deciduous stands in habitats with good soils.

Temperate Coniferous Forests

Sierra Nevada

The Sierra Nevada Range is about 100 miles long and 50 miles wide and runs southwest to northwest approximately along the border of the US American States California and Nevada. The eastern escarpment is much steeper than the western. It is home of one of the most diverse coniferous forests on Earth with an extraordinary range of habitats. Different communities are distributed in elevation belts on both sides of the range.

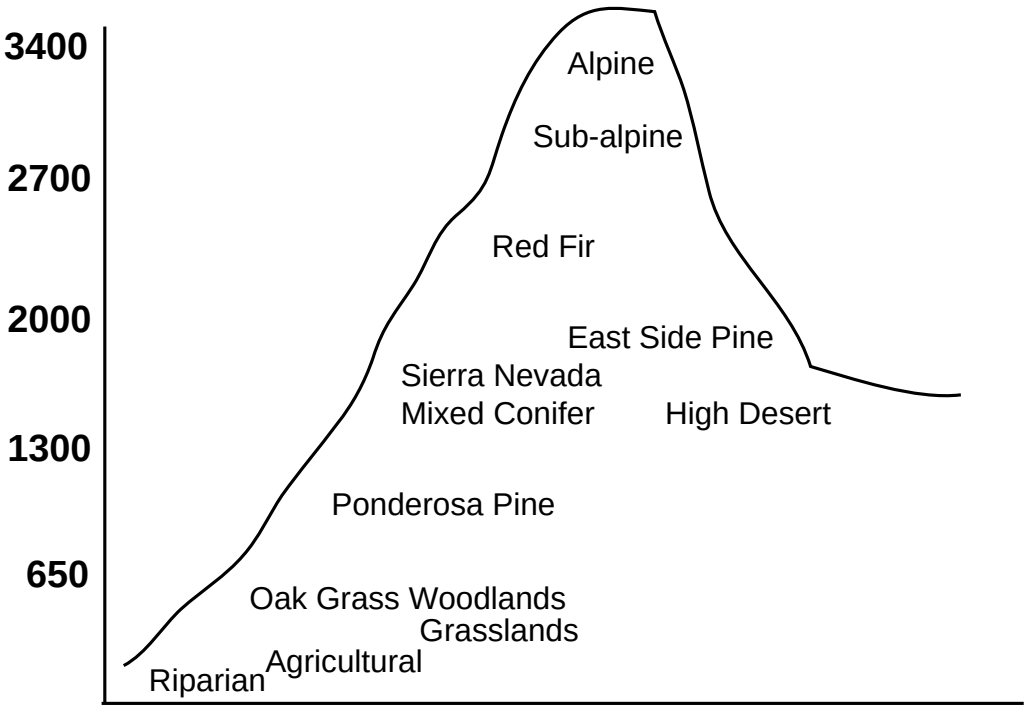

Fig. 1.3 Sierra Nevada plant communities elevations belts; heights are given in metres; changed after LeBlanc and Harris, http://www.cnr.berkeley.edu/departments/espm/extension/ECOLOGY.HTM

Klamath Siskiyou Forests

The Klamath Siskiyou forests are one of the biodiversity hotspots on Earth due to its complex terrain, geology, climate and biogeographic history. It is located in southern Oregon and northern California. The regions was not affected by quaternary glaciations and therefore provided a refuge and long periods of favourable conditions for many taxa. Climatic shifts over time made this ecoregion a transition zone of several major biota (Great Basin, Oregon Coast Range, Cascade Range, Sierra Nevada, California Central Valley). 3,500 plant species are known from this region including many endemics which only occur on single mountains. Serpentine soils are present in some spots.

Wasatch and Uinta Montane Forests

The Wasatch/Uinta ecoregion is a block of high montane habitats bordering the Great Basin in the East, reaching from southern Idaho to the isolated ranges of the Colorado Plateau. The dominant vegetation is coniferous forest with *Pinus ponderosa*, *Pseudotsuga menziesii*, *Abies lasiocarpa* and *Picea engelmann*. *Pinus flexilis* occurs rarely. The Wasatch/Uinta range is arid due to the rain shadow of the Sierra Nevada located 500 miles further west.

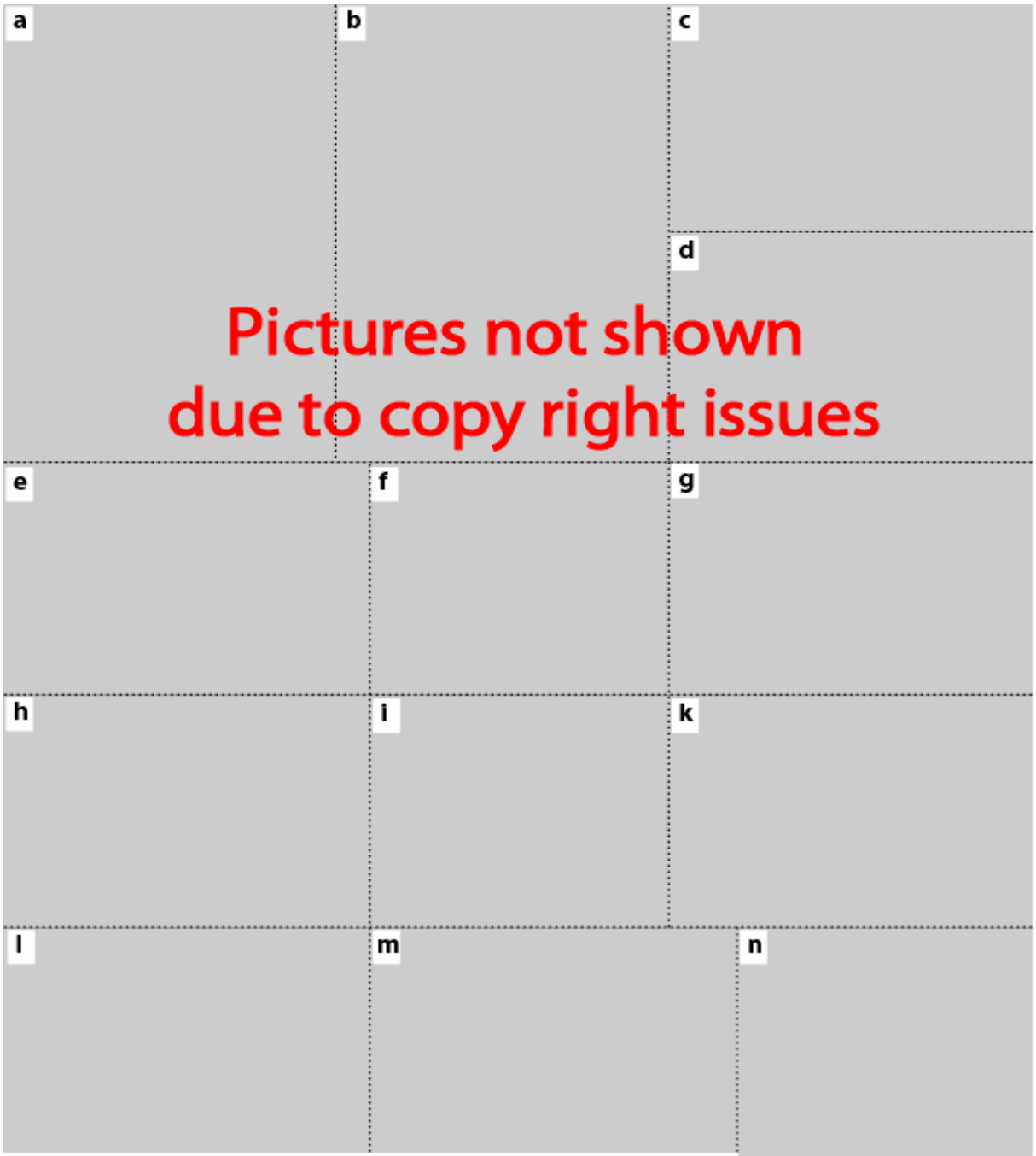

Fig. 1.4 Representative pictures of temperate ecoregions taken from www.nationalgeographic.com/wildworld. a New England Acadian Forest (S. Leslie), b Western Great Lakes Forest (A. I. Solyum), c Eastern Great Lakes Forest (the GLOBE Program), d Sierra Nevada Forests (David Olson), e Klamath-Siskiyou Forest (David Olson), f Wasatch and Uinta montane forest (John Morrison), g Colorado Rockies (Yen-Wen Lu),h North Central Rockies (David Olson), i South Central Rockies (John Morrison), k Cascade Mountains Leeward (R. Careless), l Eastern Cascades (John Morrison), m Blue Mountains (DarenHSpencer), n Canadian Aspen Forest and Parkland (D. Poulton)

Colorado Rockies Forest

The Colorado Rockies are the region in which the highest elevations of the Rocky Mountains are found. Vegetation is present in vertical zonation being a consequence of abrupt elevational gradients. The dominant vegetation type is coniferous forest along with mountain meadows, foothill grasslands, riparian woodlands and upper-treeline alpine tundra communities.

North Central Rockies Forest

The Northern Central Rockies stretch over 600 miles from north to south from western Canada into the United States. Climatic conditions vary from east to west, the west experiencing the influence of the Pacific while the east is more continental. Mean annual temperature is 2 °C higher in the western part (3.5 °C versus 5.5 °C). Glaciation left large valleys throughout the region. The dominant vegetation type is coniferous forest. Typically the tree species found in the Cascades are also found in the Northern Central Rockies. Mountain meadows, alpine habitats, foothill grasslands and riparian woodlands are also present. The vertical zonation is pronounced as a consequence of abrupt changes in elevation.

South Central Rockies Forest

The South Central Rockies cover the area western Wyoming, eastern Idaho and central Montana. A second portion is located in central and eastern Idaho south of the Clearwater River. Flatlands and mountains change abruptly resulting in a dramatic vertical zonation. The Black Hills unit of this ecoregion is the lowest in elevation but contains a distinct, floristic diversity with floral elements of the Great Basin, Eastern deciduous, Boreal, Rocky Mountain, and Southern Great Plains. The dominant vegetation type is coniferous forest. Relative to other regions within the Rocky Mountains the South Central Rockies are dry with a predominantly continental climate. Summers are short, winters are long and cold. Parts of the region are influenced by geothermal activities (Yellowstone National Park).

Cascade Mountains Leeward Forest

This ecoregion is located along the leeward side of the Cascade Mountains. A strong climatic gradient exists from the moist coastal climate to the semi-arid continental climate in the southern interior. Alpine tundra, montane forests and parklands with scattered ponderosa pine are common.

Eastern Cascades Forest

The eastern Cascades span the eastern slopes of the Cascade Mountains in Oregon and Washington meeting the Cascade Mountains Leeward in the North. The Eastern Cascades are a series of steep, rugged mountains rising up to 2,700 m and volcanic peaks reaching a hight of 4,300 m. In some areas serpentine soils are

found. The climate is in general mild. The natural vegetation is a mosaic of shrub lands, grasslands and coniferous forests. Ponderosa pine is very common but species composition of the forests varies along with environmental gradients of physical factors such as temperature and moisture.

Blue Mountains Forest

The Blue Mountains ecoregion is characterized by several basin and range areas, alluvial fans and floodplains. The relief is highly variable. Plant communities in the different ecosystems in the Blue Mountains are sagebrush-, Pinyon-juniper-, ponderosa pine-, Douglas fir-, western larch-, spruce-fir- and lodgepole pine communities as well as chaparral-mountain shrub, mountain meadows, mountain grasslands and alpine communities.

Temperate Grasslands/Savanna/Shrub

Canadian Aspen Forests and Parklands

The ecoregion is classified by a sub humid, low-boreal ecoclimate. Summers are short and warm, winters are cold and long. The region is mostly underlain by Cretaceous shale. Lakes are frequent in shallow depressions. Vegetation is characterized by *Populus tremuloides* followed in number by *Populus balsamifera* together with an understory of mixed herbs and shrubs. Picea glauca and Abies balsamea are the climax species. However, they are not frequent due to fire. Poorly drained sites are covered with sedges, willows, spruce and *Larix laricina*.

Xeric Shrublands/Deserts

Great Basin Shrub Steppe

The Great Basin is the most northerly of the four North American deserts. The Great Basin has affinities with the cold vegetation unlike the other deserts which are more associated with warm-temperate or tropical-subtropical vegetation. The topography of the Great Basin is diverse. A series of parallel ranges and their intervening valleys characterizes the landscape. The Great Basin is bordered by the central Rockies and the Colorado Plateau in the east, the Columbia Plateau in the north and the Cascades and Sierra Nevada in the west. Towards the south it meets the Mojave desert at the Colorado River drainage. Vegetation is dominated by cold temperate species like sagebrush, saltbrush and winterfat. Evolutionary tied to warmer places are genera like *Chrysothamnus*, *Coleogyne*, *Tetradymia* and *Grayia* which also occur in the Great Basin.

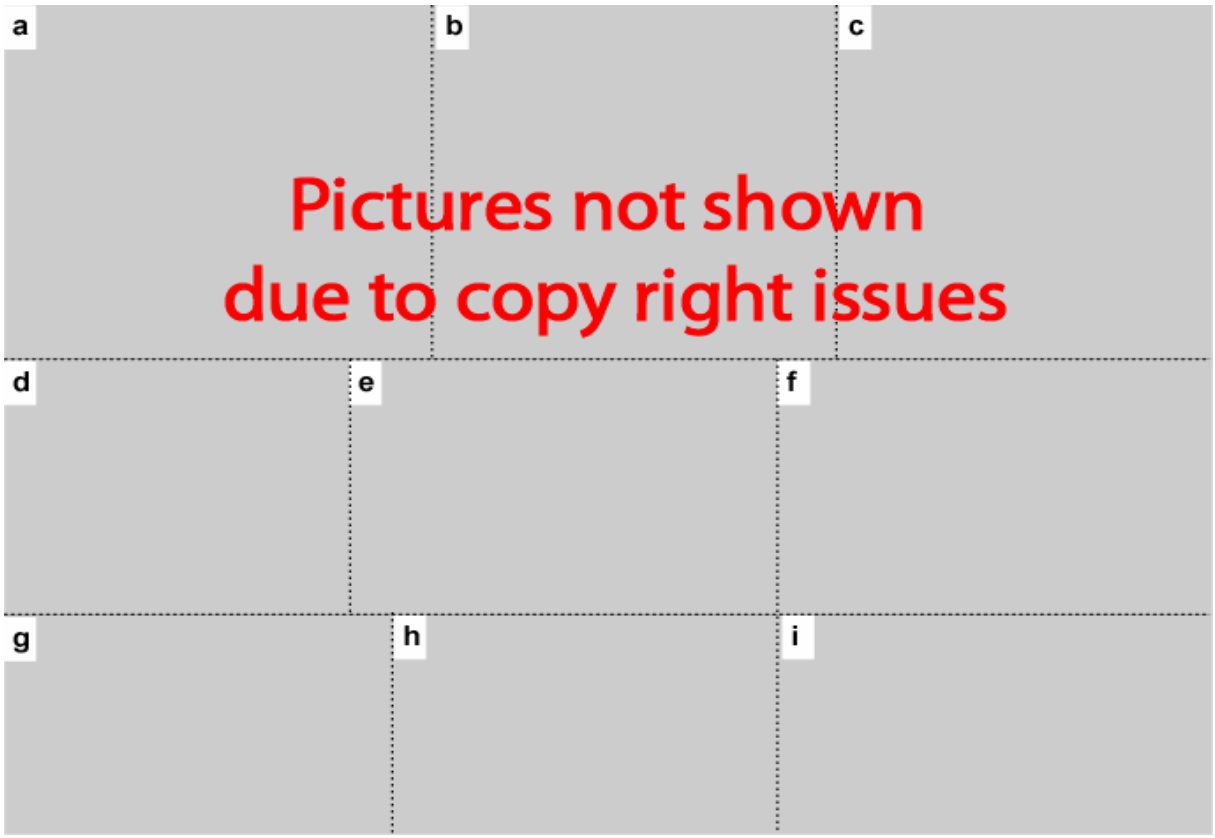

Fig. 1.5 Pictures representing dry ecoregions. Pictures all obtained from http://www.nationalgeographic.com/wildworld/, photographer given in brackets after the ecoregion name. a Great Basin (John Morrison), b Snake/Columbia Plateau (John Morrison), c Colorado Plateau (John Morrison), d Mojave Desert (John Morrison), e Chihuahua Desert (wwf-Canon, Edward Parker), f Sonora Desert (David Olson), g Wyoming Basin (USFWS), h California Montane Chaparral and Woodlands (John Morrison), i Yukon Interior Dry Forest (J. Peepre)

Snake/Columbia Shrub Steppe

The Snake/Columbia shrub steppe is a mostly arid ecoregion north of the Great Basin. Towards the east it's boundary is the Continental Divide. It is situated in a major river system. However, due to the rain shadow of the Cascade Mountains the ecoregion receives only little water. The dominant vegetation is *Artemisia* species, often associated with *Agropyron* species, *Festuca idahoensis* and other bunchgrasses. In the montaneous parts of this ecoregion juniper and bunchgrass communities are found. In some parts of montaneous areas *Pseudotsuga menziesii*, *Abies lasiocarpa* and *Populus tremuloides* occur.

Colorado Plateau Shrublands

The Colorado Plateau is an elevated, northward-tildes plateau with arid to semiarid climate. It has developed a great relief through erosion by rivers such as the Grand Canyon. The major vegetation types are woodlands dominated by *Pinus edulis* and several juniper species. Between the trees the ground is sparsely covered. The

montaneous areas are covered by ponderosa pine forests in the south and lodgepole pine and aspen forests further north. The lowest montaneous zone supports grasslands with many bare areas.

Mojave Desert

The Mojave desert is the smallest of the four North American deserts. Dominant plants of the Mojave are creosote bush (*Larrea tridentata*), brittlebush (*Atriplex polycarpa*), desert holly (*Atriplex hymenelytra*), white burrobush (*Hymenoclea salsola*), and Joshua tree (*Yucca brevifolia*). The Mojave supports numerous species of cacti. Elevations range from below sea level to 1,600 m. Most regions are between 600 and 1,200 m, this being the reason why the Mojave is referred to as a high desert.

Chihuahua Desert

The Chihuahua desert reaches from the southern United States deep into Mexico. The Chihuahua desert is in general cooler and wetter than the other North American deserts due to its higher elevation. *Larrea divarecata*, *Florensia cernua*, *Prosopis articulata* and *Acacia species* characterize the landscape. The North American Chihuahua desert has an extensive grass component as opposed to the Mexican part which supports mainly a vegetation composed of cacti, yucca and shrubs. The Chihuahua desert is of recent origin.

Sonora Desert

The Sonora Desert reaches from southeastern California into the western two thirds of southern California and into Mexico in the states of Sonora and parts of Baja California. The Sonora desert is divided into two sections characterized by the availability of water. Temperatures are relatively high in winter and summer and rainfall patterns are irregular. Lower-elevation areas are dominated by dense communities of creosote and white bursage. On slopes *Cercidium* species and ironwood (*Olneya tesota*) are found.

Wyoming Basin Shrub Steppe

The Wyoming Basin ecoregion is a high, open, arid country. It is nearly completely surrounded by montaneous ecoregions. The aridity is due to the rain shadow of the Rocky Mountains. The dominant vegetation type is sagebrush, often associated with wheatgrasses or fescue.

Boreal Forest/Taiga

Yukon Interior Dry Forests

The Yukon Interior Dry Forest lies mainly within the Yukon Territory. The climate is cold and semi-arid. The mean annual temperature is -3 °C. Elevations in this ecoregion are generally above 1,000 m. White and Black spruce form the dominant vegetation. South-facing slopes are often characterized by grassland communities. In the colder alpine regions *Dryas hookeriana*, dwarf shrubs, forbs, grasses and lichens cover the ground.

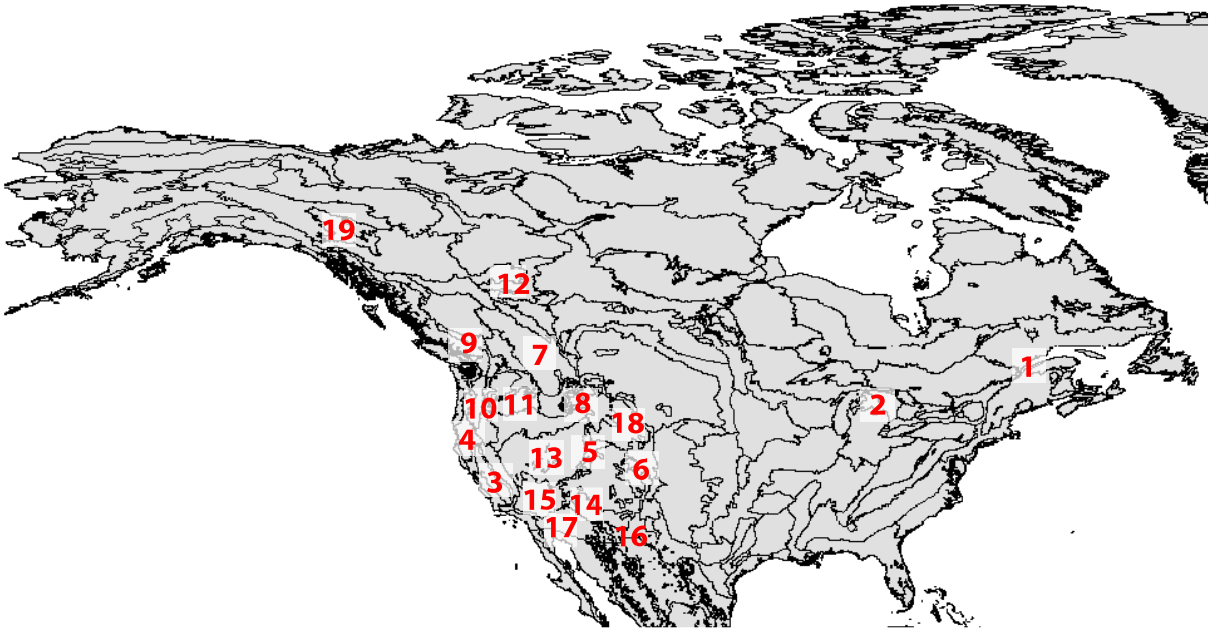

Fig. 1.6 Geographical Location of ecoregions in North America according to the wwf ecoregion map (Olson et al. 2001); 1 New England/Acadian Forest, 2 Great Lakes Forest, 3 Sierra Nevada, 4 Klamath/Siskiyou Forests, 5 Wasatch Uinta, 6 Colorado Rockies, 7 North Central Rockies, 8 South Central Rockies, 9 Cascade Mountains Leeward, 10 Eastern Cascades, 11 Blue Mountains, 12 Canadian Aspen Forest and Parkland, 13 Great Basin, 14 Colorado Plateau, 15 Mojave Desert, 16 Chihuahua Desert, 17 Sonora Desert, 18 Wyoming Basin, 19 Yukon Interior Dry Forest

1.1.4 Phylogeography in North America

Besides Europe North America is the second best studies continent in phylogeography. Western North America is highly diverse in its landscape and ecosystems and therefore species show complex distribution patterns. However, three major patterns exist which are shared by several species.

Mesic Forest Disjunction pattern - Species of mesic forests are found in the Pacific coastal and interior Rocky Mountains which are 300 km apart separated by the arid Columbia Basin. Possible explanations for this pattern are ancient vicariance with a

formerly continuous distribution of the mesic species or inland dispersal, assuming a recent colonization of the interior Rockies by mesic species (Brunsfel et al. 2001).

Cascade/Sierra pattern - It has been shown for numerous species that they reach their southern or northern distribution limit at the Sierra/Cascades transition. The reason for this may be an environmental gradient which existed during the Cenozoic, or the presence of multiple refugia (Brunsfeld et al. 2001; Calsbeek et al. 2003).

Northern Rocky Mountain pattern - A greater part of the diversity of mesic forest species occurs south of the the limits of the last Cordilleran ice-shield. The highest diversity is found in the Clearwater Range. However, some plant species from coastal ranges only occur north of the last glaciation. In general the biogeographic discontinuity in the Rocky Mountains spans two elevation belts – middle elevation (mesic-temperate species) and subalpine habitats (Cascade/Sierra species). The genetic pattern present might either be explained by multiple refugia or by recent colonization (Brunsfeld et al. 2001).

Another common pattern is the split into eastern and Western North American clades which came into secondary contact in western North America as it is found for example in several bird species (Hull and Girman, 2005; Spellman et al. 2007).
A comparison of several phylogeographic studies centred in western North America is given in table 1.1.

Table 1.1 Comparison of phylogeographic studies centred in North America

species	area under investigation	refuge area	remark	marker system	reference
Tamias ruficaudus **(Rodentia, Sciuridae)**	Northern Rocky Mountains	Clearwater Range might be one of several refugia	range fluctuations with Pleistocenic ice-ages; recolonization through Bitterroot Range and Continental Divide	mtDNA sequence data	Good and Sullivan, 2001
Cardamine constancei **(Brassicaceae)**	South and North central Rocky Mountains	"Greater Clearwater Refugium"	multi-compartmented refugium	cpDNA sequence data	Brunsfeld and Sullivan, 2005
Salix melanopsis **(Salicaceae)**	South and North central Rocky Mountains; Pacific North West	"Greater Clearwater Refugium"	no endemic haplotypes in Bitterroot Range;	cpDNA sequence data	Brunsfeld et al., 2007
Sitta carolinensis **(Aves)**	North American continent wide	refugia east and west of Great Basin in *Pinus ponderosa* stands; multible refugia possible	long-term isolated populations; divergence of northern and southern clades forced by ice-ages	mtDNA sequence data	Spellman and Klicka, 2007
Crotalus willardi obscurus **(Reptilia)**	mountain ranges along US.-México Border	-	northward migration from México and subsequent isolation on sky-islands	mtDNA and 9 microsatellites	Holycross and Douglas, 2007

species	area under investigation	refuge area	remark	marker system	reference
***Sidalcea* (Malvaceae)**	western North America	-	supports a migration from México along Sierra Nevada one one hand and Rocky Mountains on the other hand for some species	nrDNA	Andreasen and Bladwin, 2003
***Sitta pygmaea* (Aves)**	western North America	southern California (like *Pinus ponderosa*)	refugia were expected to be in south Sierra Nevada and Arizona/New Mexico	mtDNA sequence data	Spellman and Klicka, 2006
***Martes americana* (Mammalia)**	western North America	North Pacific Coast	two colonization events for pacific NW	mtDNA sequence data	Stone et al. 2002
***Dendroctonus rufipennis* (Curculionidae)**	North American continent wide	Beringia, Rockies and central eastern United States south of LGM	secondary contact in Pacific northwest	mtDNA sequence data	Maroja et al. 2007
Pinus flexilis	northwestern North America	east of the Rockies in Colorado and Kansas, numerous sites in Great Basin and Fremont Co., Colorado and Bighorn Co., Wyoming	secondary contact zone in Colorado	mtDNA sequence data	Mitton et al. 2000
***Aglenopsis aperta* (Araneae)**	western North America	east of the Rocky Mountains, between Rocky Mountains and Sierra Nevada, west of the Sierra Nevada	postglacial expansion out of eastern refuge into Rocky Mountains	mtDNA sequence data	Ayoub and Riechert, 2004
***Poecile gambeli* (Aves)**	western North America	eastern (Rocky Mountain) and western (Sierra Nevada) refuge	eastern and western clade; postglacial introgression in Mono Lake region	mtDNA sequence data	Spellman et al. 2007
***Accipiter striatus velox* (Aves)**	North American continent wide	-	distinct eastern and western group; recent expansion of eastern group	mtDNA sequence data	Hull and Girman, 2005
***Tamias amoenus* (Rodentia, Sciuridae)**	NW North America	-	phylogeography shaped by vicariance; east/west dichotomy	mtDNA sequence data	Demboski and Sullivan, 2003
***Lycaeides* (Lepidoptera)**	western North America	three refugia: central North America, east of Great Lakes, Sierra Nevada	glaciations fostered divergence, postglacial expansion lead to gene flow and reticulation	mtDNA, SSCP analysis	Nice et al., 2005
***Rana luteiventris* (Amphibia, Ranidae)**	Great Basin and adjacent mountain ranges	-	one Bonneville Basin and one Lahontan Basin clade (Nevada); one Rocky Mountain clade; no genetic structure for more recent events	mtDNA sequence data	Bos and Sites, 2001
***Batrachoseps* (Amphibia)**	California	-	large number of contact zones	mtDNA sequence data	Jockusch and Wake, 2002

species	area under investigation	refuge area	remark	marker system	reference
***Veronica alpina* (Scrophulariaceae)**	western North America	maybe northern Rocky Mountains;	three separate phylogeographic groups: Cascades, northern and southern Rocky Mountains; British Columbia as secondary contact zone	cpDNA sequence data; AFLP	Albach et al. 2006

Eastern North America has been subject to a larger number of phylogeographic studies than western North America. Phylogeographic patterns are mainly shaped by the Appalachian Mountains. Due to its north-south orientation the Appalachian Mountains did not represent an obstacle to migration into southern areas like for example the Alps in Europe. Together with the high complexity of the landscape this fact results into a multitude of phylogeographic patterns rather than the defined patterns that have been inferred for Europe (Soltis et al. 2006). Phylogeographic patterns in eastern North America have been reviewed and some recurrent patterns have also been found in eastern North America (Soltis et al. 2006). Among these were the Apalachicola River Basin discontinuity in Florida exhibited by a number of fish and turtle species and a phylogeographic split along the Tombigbee River in Alabama. A third pattern is referred to as the Appalachian Mountain discontinuity which refers to an east/west split over the mountain chain probably caused by two different glacial refugia on the opposing sites.

1.2 The Brassicaceae Family

The Brassicaceae (mustard family) comprises approximately 3500 species distributed among approximately 350 genera of which many are monotypic. Among them are well known as important crop plants such as *Brassica oleracea* or ornamental plants like *Aubrieta deltoides*. Furthermore the Brassicaceae include the most studied organism studied in plant molecular biology: the Thale Cress *Arabidopsis thaliana*. Brassicaceae are found worldwide except in the tropics and Antarctica. Their centre of diversity is the Irano-Turranian floristic region and some authors suppose an origin of the whole Brassicaceae family for this region (Hedge et al. 1976, Franzke et al. accepted).

Past taxonomic systems of the Brassicaceae followed morphological characters such as fruit or hair shape (Jahnchen, Schulz, Hayek). Traditionally the Brasicaceae have been subdivided into a tribal system. With the availability of molecular markers it was possible to reconstruct a family wide ITS (Internal Transcribed Spacer) based phylogeny. This lead to the restructuring of old tribes and also the description of new tribes such as the Boechereae (Bailey et al. 2006). However, although the resolution within the tribes was well, the relationships among the tribes was only poorly resolved. Therefore the “Brassicaceae working group” (an international group of collaborating researchers) has defined a TOP 100 list of Brassicaceae representing

the different tribes for which several nuclear markers shall be amplified and sequenced in order to obtain a better resolution in phylogenetic tree reconstruction (Marcus Koch, personal communication).

1.3 The North American Genus *Boechera* (Brassicaceae)

The genus *Boechera* comprises 110 species of which 108 are confined to the North American continent (Windham and Al-Shehbaz 2006, 2007a, 2007b). The other two species occur in Greenland (*Boechera holboellii*) and Siberia (*Boechera falcata*), respectively. *Boechera* belongs to the tribe Boechereae which also includes the genera *Nevada*, *Pennellia*, *Sandbergia* and *Cusikiella* (Bailey et al. 2006).

The plants are herbaceous and biennial to perennial with simple or branched trichomes which may be stalked or sessile. Stems can be simple or branched. Basal leaves are petiolate, simple, entire or dentate (rarely lyrat-pinnatified) while cauline leaves are sessile or rarely very short petiolate. Inflorescences are sometimes in panicles, fruiting pedicels can be erect, ascending, divaricate, or reflexed. Petals are white, pink or purple. Flower morphology follows the typical Brassicaceae flower (four sepals, four petals, two short +four long stamen and a fused pistil made up of two carpels). The ovary is superior. Fruits are siliques (Al-Shehbaz, New Flora of North America, unpublished). Examples of *Boechera* morphology are given in figure 1.7.

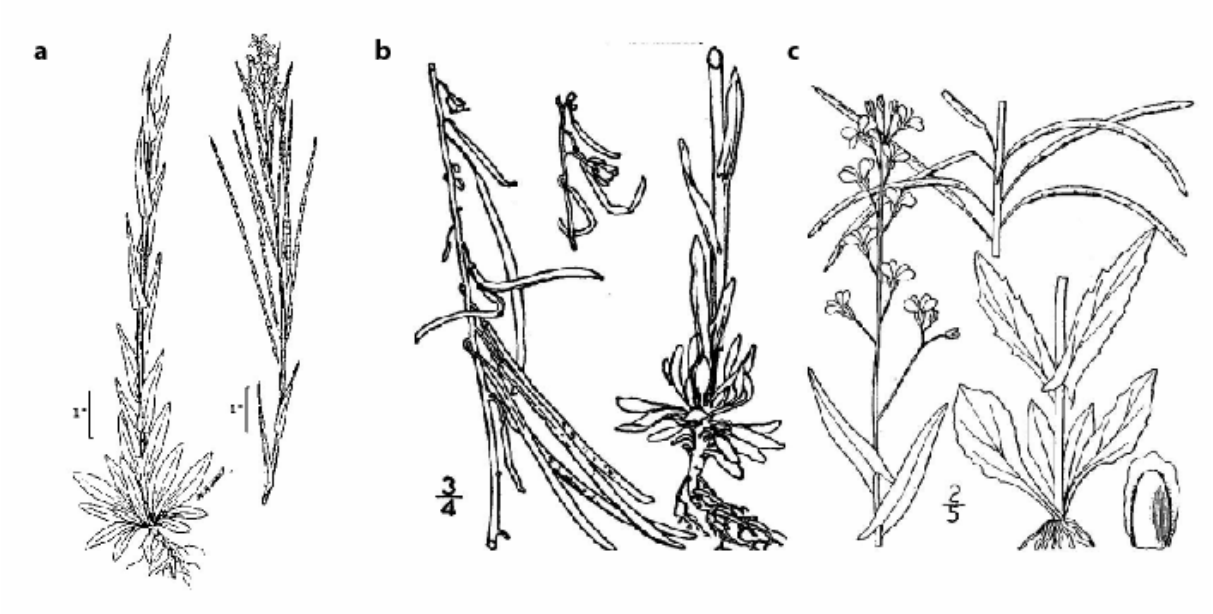

Fig. 1.7 Examples of *Boechera* morphology illustrated by a *Boechera* stricta; USDA NRCS. Wetland flora: Field office illustrated guide to plant species. USDA Natural Resources Conservation Service. Provided by NRCS National Wetland Team, Fort Worth, TX; b *Boechera holboellii* var. collinsii; Britton, N.L., and A. Brown. 1913. An illustrated flora of the northern United States, Canada and the British Possessions. Vol. 2: 183. Courtesy of Kentucky Native Plant Society; c *Boechera* laevigata; Britton, N.L., and A. Brown. 1913. An illustrated flora of the northern United States, Canada and the British Possessions. Vol. 2: 181. Courtesy of Kentucky Native Plant Society;

In the past *Boechera* used to be included into the genus *Arabis* and species number varied between 50 and 80 according to the author although *Boechera* as a genus was already established in 1976 (Löve and Löve, 1976). Later molecular data proved that a greater part of North American *Arabis* were not related to each other (Koch et al. 1999, 2000). Since then species were repeatedly transferred into the genus *Boechera*, named after the Danish botanist Tyge Böcher who cytogenetically studied the plants in great detail (Böcher, 1951). The most recent description of the whole genus was done for the new edition of the Flora of North America (Al-Shehbaz, New Flora of North America, unpublished), giving a more extensive and complete description of the genus than the series of papers published by Windham and Al-Shehbaz (2006, 2007a, 2007b).
In *Boechera* the base chromosome number is x=7. However, plants do not only occur on the diploid level. Aneuploidy is a common phenomenon not only represented by the many triploids but also in chromosome numbers deviating from a multiple of seven. Some diploid plants and the aneuploids reproduce apomictically (asexual reproduction via seeds). In the new description of *Boechera* sexually and asexually reproducing plants were separated into different species. Hence 72 species are sexually reproducing while 38 species which are assumed to be of hybrid origin are asexually reproducing.
Boechera lives in a wide range of different habitats reaching from dry desert and shrub steppe to forest and alpine habitats.

1.4 Apomixis in *Boechera*

Boechera may reproduce sexually or asexually via apomixis. Cytogenetic studies showed that apomicts contain chromosomes from *Boechera stricta* and *Boechera holboellii* relatives (sensu Rollins, 1993) and my therefore be assumed to be of hybrid origin (Kantama et al. 2007). Morphological evidence also supports this view (Windham and Al Shehbaz, 2006. 2007a, 2007b).
Often apomixis is associated with high ploidy levels. However, in *Boechera* apomicts can be diploid, triploid or aneuploid (Dobeš et al. 2006). Apomixis in *Boechera* is facultative and sexual and asexual reproduction may occur in the same individual (Schranz et al. 2005). Hence genomes can be recombined leading to different numbers of parental chromosomes in the hybrid offspring (e.g. four B*. stricta* chromosomes and ten *B. holboellii* related chromosomes; Kantama et al. 2007). Microsporogenesis in *Boechera* apomicts is disturbed and leads to a number of malformed pollen grains along with a number of spherical pollen grains opposed to the elongated ones produced by sexuals (Voigt et al. 2007). Pollen size was also taken as a criterion to distinguish sexual and asexual *Boechera* species for the new species circumscriptions (Windham and Al Shehbaz, 2006).
It was hypothesized that *Boechera* apomicts are characterized by a B chromosome (Sharbel et al. 2004, Sharbel et al. 2005). B chromosomes are extra chromosomes

which are inherited independently, are dispensable and do not pair with other chromosomes during meiosis (Jones and Houben, 2003). However, cytogenetic studies could show that 2n and 2n + 1 apomicts are rather characterized by the presence of a heterochromatic chromosome and in the cases of the 2n + 1 apomicts a deletion chromosome. According to their most prominent features the chromosomes were named *Het* and *Del* (Kantama et al. 2007). Meiotic studies suggest that the Het chromosome pairs regularly with another chromosome while the Del chromosome forms a heteromorphic trivalent. Since *Boechera* chromosomes are very small (3-6 µm) and uniform in shape and size (except the *Het* and *Del*) it is unknown to which other chromosome/chromosome part *Het* and *Del* are homologous. Since the *Het* chromosome is present only in apomicts it is tempting to hypothesize that it plays a role in expressing apomixis. However, nothing which would support this fact has been proven so far.

1.5 Aims of the Dissertation

The aim of the research conducted in the framework of this thesis was to unravel the phylogeographic and evolutionary history for as many members of the genus *Boechera* as possible by using cpDNA data following a study which had been based on *Boechera divaricarpa* (sensu Rollins, 1993), *Boechera stricta* and *Boechera holboellii* (sensu Rollins, 1993) (Dobeš et al. 2004). Centres of genetic diversity should be revealed for the whole genus and should be compared with species diversity data.

A second aim was to nrDNA ITS gene pool and to reconstruct a phylogeny for the genus by using nrDNA and nuclear single copy gene markers (introns from genes neighbouring the loci encoding the vrn1 and ELF3 proteins; primers according to Eric Schranz, Thomas Mitchell-Olds and Bao-Hua Song).

To get an insight into the genomic structure of apomicts BAC-FISH analyses were carried out in order to identify translocations and the origin of the *Het* and *Del* chromosome (project part funded by the DAAD).

1.6 Literature cited

Albach, D. C., Schönswetter, P., Tribsch, A. 2006. Comparative phylogeography of the *Veronica alpina* complex in Europe and North America. Molecular Ecology 15: 3269–3286.

Al-Shehbaz, I. A. submitted. *Boechera* A. Löwe and D. Löwe. In: Flora North America.

Andreasen K., Baldwin, B. G. 2003. Reexamination of relationships, habital evolution, and phylogeography of checker mallows (*Sidalcea*; Malvaceae) based on molecular phylogenetic data. American Journal of Botany 90: 436–444.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., Saunders, N. C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Reviews in Ecology and Systematics 18: 489–522.

Avise, J. C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7: 371–379.

Ayoub, N. A., Riechert, S. 2004. Molecular evidence for Pleistocene glacial cycles driving diversification of a North American desert spider, *Agelenopsis aperta*. Molecular Ecology 13: 3453–3465.

Bailey, C. D., Koch, M. A., Mayer, M., Mummenhoff, K., O'Kane, S. L., Warwick, S. I., Windham, M. D., Al-Shehbaz, I. A. 2006. A Global nrDNA ITS Phylogeny of the Brassicaceae. Molecular Biology and Evolution 23: 2142–2160.

Birky, C. W. Jr. 1995. Uniparental Inheritance of mitochondrial and chloroplast genes: Mechanisms and Evolution Proceedings of the National Academy of Sciences 92: 11331–11338.

Böcher, T. W. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Kong Danske Vidensk Selsk, 6: 1–59.

Bos, D. H., Sites, J. W. Jr. 2001. Phylogeography and conservation genetics of the Columbia spotted frog (*Rana luteiventris*; Amphibia, Ranidae). Molecular Ecology 10: 1499-1513.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., Soltis, P. S. 2001. Comparative phylogeography of north-western North America: a synthesis. *in*: Integrating ecology and evolution in a spatial context. Silvertown J.; Antonovics J. Blackwell Science LTD. ISBN 0-632-05824-2.

Brunsfeld, S. J., Miller, T. R., Carstens, B. C., 2007. Insights into the Biogeography of the Pacific Northwest North America: Evidence from the Phylogeography of *Salix melanopsis*. Systematic Botany 32: 129–139.

Brunsfeld, S. J., Sullivan, J. 2005. A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conservation Genetics 6: 895–904.

Calsbeek, R., Thompson, J. N., Richardson, J.E. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Molecular Ecology 12: 1021–1029.

Delcourt P. A., Delcourt H. R. 1981. Vegetation maps for eastern North America: 40,000 yr BP to the present. *Geobotany II*. Ed; R.C. Romans. Plenum.

Lisiecki, L.E., 2005. "Ages of MIS boundaries." LR04 Benthic Stack Boston University, Boston, MA

Dembiski, J. R., Sullivan, J. 2003. Extensive mtDNA variation within the yellow-pine chipmunk, Tamias amoenus (Rodentia: Sciuridae), and phylogeographic inferences for western North America. Molecular Phylogenetics and Evolution 26: 389–408.

Dobeš, C. H., Mitchell-Olds, T., Koch, M. A., 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* X *divaricarpa*, and *A. holboellii* (Brassicaceae). Molecular Ecology 13: 349–370.

Dobeš, C., Koch, M. A., Sharbel, T. F. 2006. Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Annals Miss. Bot. Gard. 93: 517–533.

Franzke, A., German, D., Al-Shehbaz, I. A.,Mummenhoff, K. accepted. *Arabidopsis* family ties: Molecular phylogeny and age estimates in the Brassicaceae. TAXON

Good, J. M., Sullivan, J., 2001. Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. Molecular Ecology 10: 2683–2695.

Hedge, I. C. 1976. A systematic and geographical survey of the Old World Cruciferae. In The biology and chemistry of the Cruciferae (eds. Vaughan JG, Mac Leod AJ, Jones BMG), pp. 1–45. Academic Press, London.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transaction of the Royal Society of London B. Biological Sciences 359:183–95.

Holycross, A. T., and Douglas, M. E. 2007. Geographic isolation, genetic divergence, and ecological non-exchangeability define ESUs in a threatened sky-island rattlesnake. Biological Conservation 134: 142–154.

Hull, J. M., Girman, D. J. 2005. Effects of Holocene climate change on the historical demography of migrating sharp-shinned hawks (*Accipiter striatus velox*) in North America. Molecular ecology 14: 159–170.

Jackson S. T., Overpeck J. T., Webb T. III, Keattch S. E., Anderson K. H. 1997. Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews* v.16 p.1–70.

James, L. A., Harbor, J., Fabel, D., Dahms, D., Elmore, D. 2002. Late Pleistocene Glaciations in the Northwestern Sierra Nevada, California. Quaternary Reserach 57: 409–419.

Jokusch, E., Wake, D. B. 2002. Falling apart and merging: diversification of slender salamanders (Plethodontidae: *Batrachoseps*) in the American West. Biological Journal of the Linnean Society 76: 361–391.

Jones, N., Houben, A. 2003. B chromosomes in plants: escapees from the A chromosome genome? TRENDS in Plant Science 8: 417–423.

Kantama, L., Sharbel, T. F., Schranz, M. E., Mitchell-Olds, T., de Vries, S., de Jong, H. 2007. Diploid apomicts of the *Boechera holboellii* complex show large scale chromosome substitutions and different aberrant chromosomes. Proceedings of the National Academy of Sciences 104: 14026–14031.

Koch, M., Bishop, J., Mitchell-Olds, T. 1999. Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biology: 1: 529–537.

Koch, M., Haubold, B., Mitchell-Olds, T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, *Arabis* and related genera. Molecular Biology and Evolution. 17: 1483–1498.

Lang, G. 1994 in Quartäre Vegetationsgeschichte Europas, Gustav Fischer Verlag, Jena. ISBN 3-334-60405-5.

Lisiecki, L. E., Raymo, M. E. 2005. "A Pliocene-Pleistocene stack of 57 globally distributed benthic d18O records. Paleoceanography." vol. 20, PA1003, doi:10.1029/2004PA001071

Löve, A., Löve, D. 1976. Botaniska Notiser 128: 513.

Maroja, L. S., Bogdanowicz, S. M., Wallin, K. F., Raffa, K. F., Harrison, R. G. 2007. Phylogeography of spruce beetles (*Dendroctonus rufipennis* Kirby) (Curculionidae: Scolytinae) in North America. Molecular Ecology 16, 2560–2573.

Mitton, J. B., Kreiser, B. R., Latta, R. G. 2000. Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. Molecular Ecology 9, 91–97.

Nice, C. C., Anthony, N., Gelembiuk, G., Ratermans, D., Ffrench-Constant, R. 2005. The history and geography of diversification within the butterfly genus *Lycaeides* in North America. Molecular Ecology 14, 1741–1754.

Olson, D. M., Dinerstein, E., Wikramanayke, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001. Terrestrial ecoregions of the world: New map of life on earth. Bioscience 51: 933–938.

Ricketts, T. H., Loucks, C. J., Dinerstein, E. 1999. Terrestrial Ecoregions of North America: A conservation Assessment. Island Press.

Richard B. Waitt, Jr. 1985. Case for periodic, colossal jökulhlaups from Pleistocene glacial Lake Missoula. Geological Society of America Bulletin 96: 1271–1286.

Ritchie J. C., Cwynar L. C. 1982. The late Quaternary vegetation of the North Yukon, *in*: *Palaeoecology of Beringia*. ed.s Hopkins, D. M., Matthews, J. V., Schweger, C. E., Young, S. B. Academic press, London p.113–126.

Rollins, R. C. 1993. The Cruciferae of Continental North America. Stanford University Press, Stanford.

Schranz, M. E., Dobeš, C. H., Koch, M. A., Mitchell-Olds, T. 2005. Sexual reproduction, hybridization, apomixis and polyploidization in the genus *Boechera* (Brassicaceae). American Journal of Botany 92: 1797–1810.

Seibold, E., Seibold, I. 2005. Milankovitch's Strahlenkurve und deren geologische Deutung - Anfänge in Deutschland. International Journal of Earth Sciences 3: 495–503.

Sharbel, T.F., Voigt, M.-L., Mitchell-Olds, T., Kantama, L., de Jong, H. (2004) Is the aneuploid chromosome in apomictic *Boechera holboellii* a genuine B chromosome? Cytogenetic Genome Research 106: 173–183.

Sharbel, T. F., Mitchell-Olds, T., Dobeš, C., Kantama, L., de Jong, H. 2005. Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic Genome Research 109: 283–292.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., Soltis, P. S. 2006. Comparative phylogeography of unglaciated eastern North America. *Molec. Ecol*. 15: 4261–4293.

Spellmann, G. M., Klicka, J. 2006. Testing hypotheses of Pleistocene population history using coalescent simulations: phylogeography of the pygmy nuthatch (*Sitta pygmaea*). Proceedings of the Royal Society B. 273: 3057–3063.

Spellmann, G. M., Klicka, J. 2007. Phylogeography of the white-breasted nuthatch (*Sitta carolinensis*): Diversification in North American pine and oak woodlands. Molecular Ecology 16: 1729–1740.

Spellmann, G. M., Riddle, B., Klicka, J. 2007. Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to Quaternary climate change. Molecular Ecology 16: 1055–1068.

Stiff, B. J., Hansel, A. K. 2004. *Quaternary glaciations in Illinois.* in Ehlers, J., and P.L. Gibbard, eds., pp. 71-82, *Quaternary Glaciations: Extent and Chronology 2: Part II North America*, Elsevier, Amsterdam. ISBN 0-444-51462-7.

Stone, K. D., Flynn, R. W., Cook, J. A. 2002. Post-glacial colonization of northwestern North America by the forest-associated American marten (*Martes americana*, mammalia: Carnivora: Mustelidae). Molecular Ecology 11: 2049-2063.

Thompson, R. S., Whitlock, C., Bartlein, P. J., Harrison, S. P., Spaulding, W. G. 1993. Climatic changes in the western United States since 18,000 yr. B.P. p.468-513. In; *Global Climates Since the Last Glacial Maximum*. ed.s; Wright H.E. Jr., Kutzbach J.E., Webb T. III, Ruddiman W.F., Street-Perrott F.A. & Bartlein P.J.: University of Minnesota Press, Minneapolis.

Voigt, M.-L., Melzer, M., Rutten, T., Mitchell-Olds, T., Sharbel, T. F. 2007. Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. Pp. 235--258 in: Hörandl, E., Grossniklaus, U., can Dijk, P., Sharbel, T.F. (eds), *Apomixis: Evolution, Mechanisms and Perspectives*. International Association for Plant Taxonomy, Koeltz Scientific Books.

Whitlock C. & Bartlein P.J. 1997. Vegetation and climate change in northwest America during the past 125 kyr. *Nature*, v.388 p.57–61.

Windham, M. D., Al-Shehbaz, I. A. 2006. New and noteworthy species of *Boechera* (Brassicaceae) I: sexual diploids. Harvard Papers in Botany 11: 61–68.

Windham, M. D., Al-Shehbaz, I. A. 2007a. New and Noteworthy Species of *Boechera* (Brassicaceae) II: Apomictic Hybrids. Harvard Papers in Botany 11: 257–274.

Windham, M. D., Al-Shehbaz, I. A., 2007b. New and Noteworthy Species of *Boechera* (Brassicaceae) III: Additional Diploids and Apomictic Hybrids. Harvard Papers in Botany 12: 251–274.

<u>web-information:</u>

Adams J.M. & Faure H. 1997. (ed.s), QEN members. Review and Atlas of Palaeovegetation: Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Oak Ridge National Laboratory, TN, USA. http://www.esd.ornl.gov/ern/qen/adams1.html Retrieved in September 2008

Brief geologic history, Rocky Mountain National Park http://www.nature.nps.gov /geology/parks/romo/index.cfm#geology Retrieved in September 2008

Ice Age Floods, From: U.S. National Park Service Website Retrieved September http://vulcan.wr.usgs.gov/Glossary/Glaciers/IceSheets/description_ice_sheets.html

Le Blanc and Harris, http://www.cnr.berkeley.edu/departments/espm/extension/ ECOLOGY.HTM

Damery, 2004 Formation of the Great Lakes.http://www.emporia.edu /earthsci/student/damery1/gl_form.html#Pre-Wisconsin_Drainage Retrieved on 16th October 2008

2. Molecular Marker Based Studies of the North American Genus *Boechera* (Brassicaceae)

2.1 cpDNA Gene pool Analyses

Manuscript submitted to MPE; supplementary data are included on the DVD in the back of the dissertation.

PHYLOGEOGRAPHIC STRUCTURE OF THE CHLOROPLAST DNA GENE POOL IN NORTH AMERICAN *BOECHERA* – A GENUS AND CONTINENTAL WIDE PERSPECTIVE

Christiane Kiefer[1], Christoph Dobeš[1,3], Timothy F. Sharbel[2] and Marcus A. Koch[1]

[1] Heidelberg Institute of Plant Science, Biodiversity and Plant Systematics, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany

[2] Apomixis Research Group, Dept. of Cytogenetics and Genome Analysis, Leibnitz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany

[3] Present address: Department of Pharmacognosy, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

keywords: *Boechera* / Brassicaceae / cpDNA / phylogeography / hybridisation

Corresponding author:
Marcus A. Koch
Heidelberg Institute of Plant Sciences, Department Biodiversity and Plant Systematics, University of Heidelberg, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany.
Phone: +49-6221-54-4655
FAX: +49-6221-54-5508
eMAIL: mkoch@hip.uni-heidelberg.de

Running Title: cpDNA and species diversity analyses in *Boechera*

Abstract

Continental wide phylogeographic studies of plants in North America are rare. In our study we examined the phylogeographic history of *Boechera* (Brassicaceae) on a continental wide scale testing if it is possible to do an analysis for 57 of the currently accepted taxa simultaneously. We detected a large amount of haplotype sharing indicating recurrent hybridisation on the one hand and non-differentiation of haplotypes since speciation on the other hand. Hence, the chloroplast gene pool in *Boechera* predates speciation and therefore justifies the simultaneous analysis of a large number of taxa. Unrelated from taxon identity we can show that the evolutionary lineages detected have a different phylogeographic history in terms of glacial refugia and recently recolonized areas.

Introduction

Phylogeography examines the spreading of taxa in space and time by relating molecular genetic analyses to geography. The term phylogeography was introduced only about 20 years ago (Avise et al. 1987) and since then a multitude of studies investigated migration patterns of various plant and animal taxa (Avise 1998). Many of these studies focused on the evolutionary history of European and North American taxa during the Pleistocene, and the later introduced field of comparative phylogeography enabled researchers to reveal common patterns and trends. The distribution of species as we find it today is not only shaped by habitat preferences but also by the glaciation cycles with alternating cool and warm periods during the Quaternary which forced plants and animals to retreat into mostly southern located glacial refugia during colder periods. Warm periods allowed the taxa to recolonize their original distribution ranges again before they had to retreat into their refugia again during the next glacial cycle. However, plants adapted to colder habitats could also migrate and colonize new areas during periods in which temperatures decreased again in the changing climate and changing landscape which continuously provided new geographically defined corridors. The later fact might have facilitated that not only migration, but also speciation was promoted (Jordon-Thaden and Koch, accepted). These quaternary migrations left their footprints until today expressed in the geographical distribution of DNA-based polymorphisms, a phenomenon often described as “northern purity verses southern richness” (Hewitt, 2001).

Phylogeographic patterns in Europe were largely influenced by the Alps as an east-west oriented barrier for southwards migration and the Pyrenees as a barrier between the Iberian Peninsula and central Europe. Those two major obstacles lead to clear phylogeographic patterns in several plant and animal taxa eg. refuge areas close to the eastern margin of the Alps or on the Balkans or differentiation of Iberian from central European populations (Hewitt, 2001; Taberlet et al. 1998).

Phylogeographic patterns in North America are less clear partly due to the north-south orientation of North American mountain chains which meant that plants and animals could migrate along the mountain chains into southern regions without having to cross them. However, a major refuge area for several plant and animal taxa was found in the southern central Rocky Mountains where the canyons were just deep enough to maintain warmer temperatures suitable for survival (Ayoub and Riechert, 2004; Brunsfeld and Sullivan, 2005; Mitton et al. 2000). For other plant and animal taxa glacial refugia were found in the Colorado Rockies and the eastern Great Basin (Limber pine, Mitton et al. 2000) and the eastern Great Lakes (*Lycaeides*, Nice et al. 2005).

Most phylogeographic studies focus on the migration of a single species. Often only small study areas are covered due to sampling of parts of the distribution ranges only or a narrow distribution of the species. However, in order to investigate large scale

phylogeographic patterns it is necessary to study a system with a wide distribution range (e.g. Koch and Matschinger, 2007; Jakob and Blattner, 2006). One of the few continental wide phylogeographic studies in North America used three members of the genus *Boechera* (Brassicaceae): the former *Boechera holboellii* sensu Rollins (Rollins, 1993), *Boechera divaricarpa* sensu Rollins (Rollins, 1993) and *Boechera stricta*, respectively (Dobeš et al. 2004a). This study suggested the Rocky Mountains as a primary centre of genetic diversity and the Sierra Nevada as a secondary centre of genetic diversity (Dobeš et al. 2004a). Those three species were taken as representative for the whole genus since they have the widest distribution range of all *Boechera* species.
In the herein presented study we test if it is possible to perform a genus-wide phylogeographic study of *Boechera* by including 57 taxa which represent half of the currently accepted 110 species (Al-Shehbaz, unpublished). If it is possible to do a genus-wide phylogeographic analysis *Boechera* as a monophyletic group with a continental-wide distribution may be an excellent model system for examining large-scale phylogeographic patterns on the North American continent. Hence we would like to compare patterns revealed by the large-scale study to other smaller studies and see to what extent the results are congruent.
Furthermore genetic and species diversity should be compared for revealing to which extent they are congruent. Congruence of the centres of genetic and species diversity could be interpreted as long term stable populations or as regions subject to multiple colonization in which speciation took place over a long time. Incongruencies on the other hand could be interpreted as either more recent speciation events since which no divergence of cpDNA haplotypes has happened yet or as populations in which speciation had already happened and which then went through bottlenecks and lost part of their genetic diversity.

Boechera together with seven other genera belongs to the tribe Boechereae (Bailey et al. 2006). The Boechereae are almost exclusively found in North America (one species in Siberia and Greenland only, respectively). According to the most recent taxonomic concept *Boechera* comprises 110 species. Crossing experiments (Roy, 1995; Schranz et al. 2005) and studies of nrDNA ITS variation (Koch et al. 2003) showed the high potential for hybridisation within the various members of the genus. Hybrids are often triploid and reproduce asexually by apomixis (Dobeš et al. 2006). Until the complete revision of *Boechera* for the new Flora of North America (Al-Shehbaz, unpublished; Windham and Al-Shehbaz 2006, 2007a, 2007b) triploid hybrids were lumped together with the diploid taxa to which they were most similar. Taxonomically they were often treated as a variety. Today the triploid apomicts are separated from the diploid sexuals and are described as distinct species. Hence the 110 recognized species contain 72 sexually reproducing diploid species and 38 triploid apomictic species. These *Boechera* apomicts have been investigated in

various studies focusing on genome evolution (Kantama et al. 2007) or addressing the phenomenon of apomixis (Voigt et al. 2007). Here it is still unclear if apomixis is a driving evolutionary agent for speciation or simply the result of hybridization and genomic/genetic stabilization. Apomixis/polyploidy seems to have arisen repeatedly from a sexual background (Sharbel et al. 2001; Dobeš et al. 2004b) and has been extensively described for the former "*Boechera holboellii* sensu Rollins" (Rollins 1993), an almost artificial taxon (Al-Shebaz pers. Comm.). Hence, inhere we also look for apomictic lineages within *Boechera* by reconstructing networks and phylogenetic trees based on chloroplast DNA marker sequences *trnL-F* and *rpo*C1 including more than 1300 accessions collected from herbarium specimens representing 57 taxa.

Materials and Methods

PLANT MATERIAL

Leaf material was obtained from herbarium specimens from GH, MO and DAO and from collections of Thomas Mitchell-Olds (Duke University, North Carolina, USA). Corresponding accession details are listed in supplement table 1, the distribution of the samples is given in figure 7 (online material). Herein we analysed in total 1286 accessions.

DNA EXTRACTION AND SEQUENCING

Total DNA extraction and PCR reactions were done as described in Dobeš et al. (2004) with some minor changes. Prior to sequencing PCR products were purified with the NucleoFast Kit (Macherey & Nagel). Cycle sequencing was done in our lab or at the Genome Centre, IPK Gatersleben, using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) and running a MegaBASE 500 sequencer. Sequencing was done in forward and reverse reaction.

DNA MARKER SELECTION

We used various DNA regions from the chloroplast genome to reconstruct maternal phylogenetic lineages allowing also to reconstruct past migration history as the chloroplast genome is inherited maternally in the Brassicaceae. The trnLF region was analysed to characterize cpDNA haploytypes for further assessment of genetic variation and diversity. The same DNA region was also used to reconstruct haplotype networks and infer phylogenetic hypothesis. In order to increase significance of networks and phylogenetic trees with its major evolutionary lineages we also added sequence data from the rpoC1 intron to a subset of trnLF haplotypes.

ALIGNMENTS AND HAPLOTYPE DEFINITION

The forward and reverse sequences were aligned and trimmed. The *trn*L intron and the *trn*LF intergenic spacer regions were assembled into one sequence and missing bases at the 3' and 5' prime ends were substituted with N. The alignment was done manually using the program GenDoc according to the alignment published by (Dobeš et al. 2004a). Haplotypes were defined by running the program TCS 1.21 (Clement et al. 2000). Gaps were set as a 5th state and the connection limit to 95%. In *Boechera* (as in other Brassicaceae genera e.g. *Arabidopsis*, *Cardaminopsis*, refer to (Koch et al. 2007a, 2007b)). *trn*F pseudogenes are present in the *trn*L-*trn*F intergenic spacer. In past studies in our group (Dobeš et al. 2004a, Schmickl et al. 2008) it was shown that pseudogene copies were gained and lost independently several times across the phylogenetic tree. As the mechanism through which those pseudogenes arise and multiply or are deleted is unknown, we omitted them from the analysis. Neglecting the pseudogenes lead to the collapse of some haplotypes into single nodes in the network analysis. Those haplotypes are hereafter referred to as suprahaplotypes. Twenty suprahaplotypes (S1 to S20) were defined (see supplement table 3). For the *rpo*C1 intron forward and reverse sequences were also assembled and trimmed and subsequently aligned manually. For a combined analysis *rpo*C1 and *trn*LF sequence data of the same individual were fused into a combined alignment to obtain more variable sites for the analysis.

DNA-BASED PHYLOGENETIC RECONSTRUCTIONS AND NETWORK ANALYSIS

Phylogenetic reconstruction based on (1) the *trn*LF dataset and (2) a combined dataset of a subset of the *trn*LF dataset together with *rpo*C1 sequences was done by running a heuristic search with parsimony as optimality criterion in PAUP4.0beta (Swofford 2001) under default parameters. The initial max trees was set to 1000. Trees were rooted with *Halimolobus perplexa* as outgroup. The *trn*F pseudogene region (see Koch et al. 2007a, 2007b) spanning bases 837 to 1001 in the alignment was excluded in both analyses. A strict consensus tree was calculated from the 1000 shortest trees. For confirmation of the tree structures a bootstrap analysis was done. Number of bootstrap replicates was set to 1000 and the optimality criterion was parsimony. The search was also done with default parameters.

Network reconstruction was done by running the program TCS1.21 (Clement et al. 2000). Two analyses were carried out. In the first analysis the *trn*LF dataset was analysed alone. First the alignment was split into three sub-alignments according to the parsimony analysis. Then the pseudogene region was excluded and the connection limit was set to 95%. After network reconstruction the sub-networks were rejoined according to the parsimony analysis. In the second analysis a subset of the *trn*LF dataset was combined with *rpo*C1 sequences in order to get a higher resolution and to confirm the backbone of the network. *trn*F pseudogenes were again excluded. In the network analysis the connection limit was set to six steps as more steps lead

to inaccuracies. Haplotypes that were excluded from the network due to the lower connection limit were fused to their most likely position according to the phylogenetic tree. Those added connections are shown as dotted lines.
For the identification of species specific lineages suprahaplotype identity and species identity were entered in a table showing suprahaplotype identity, species identity and percentage of suprahaplotype per species.

GENETIC AND SPECIES DIVERSITY
All geographical information available from the annotation on the herbarium vouchers was entered into a BioOffice 2.0.4 (Biogis Consulting) database. Missing geographical coordinates were added according to the descriptions on the herbarium vouchers. Using BioOffice, haplotypes were plotted into the WWF ecoregions map (Olson et al. 2001), and distribution ranges of haplotype lineages were later drawn by hand on a satellite picture (Photo Courtesy of NASA).
For diversity statistics the dataset was divided into subsets according to ecoregions as defined in the WWF ecoregions map version (Olson et al. 2001).
Genetic diversity was calculated as gene diversity and nucleotide diversity as implemented in Arlequin vers. 2.0 (Excoffier et al. 2005) separately for ecoregions and lineages. In cases where the number of accessions per region was below eight, two adjacent regions with similar ecosystems were analysed together. Phylogeogographic inferences were made by applying the concept of comparing gene diversity and nucleotide diversity according to Avise (2000). Scenarios were derived for all ecoregions where a sufficient number of samples was present.
Following the taxonomic concept of Windham and Al-Shehbaz and the distribution ranges of species as given in their publications (Windham and Al-Shehbaz 2006, 2007a, 2007b) species number per state was recorded. Species numbers where colour coded and states were shaded according to the colouring scheme. Species diversity was calculated with and without apomictic species.

Results

CALIFORNIA AND NEVADA ARE THE ALPHA BIODIVERSITY HOTSPOT FOR *BOECHERA*

According to the current taxonomic concept (Al-Shehbaz, submitted) for *Boechera*, 110 species are accepted of which 72 are sexually reproducing and 38 are apomictic. Fifty-seven species are found in California, thus making it the species biodiversity hotspot for *Boechera*. Fewer species have been reported from other states, including Nevada (46), Oregon (32),Utah (30), Idaho (26) and Wyoming (23). Between 10 and 20 species were found in Colorado, Washington, Montana and Utah. Species diversity decreased to the north, south and east from California (Figure 1a).

Omitting the apomictic species from the analysis did not change the relative species diversity among the states (data not shown).

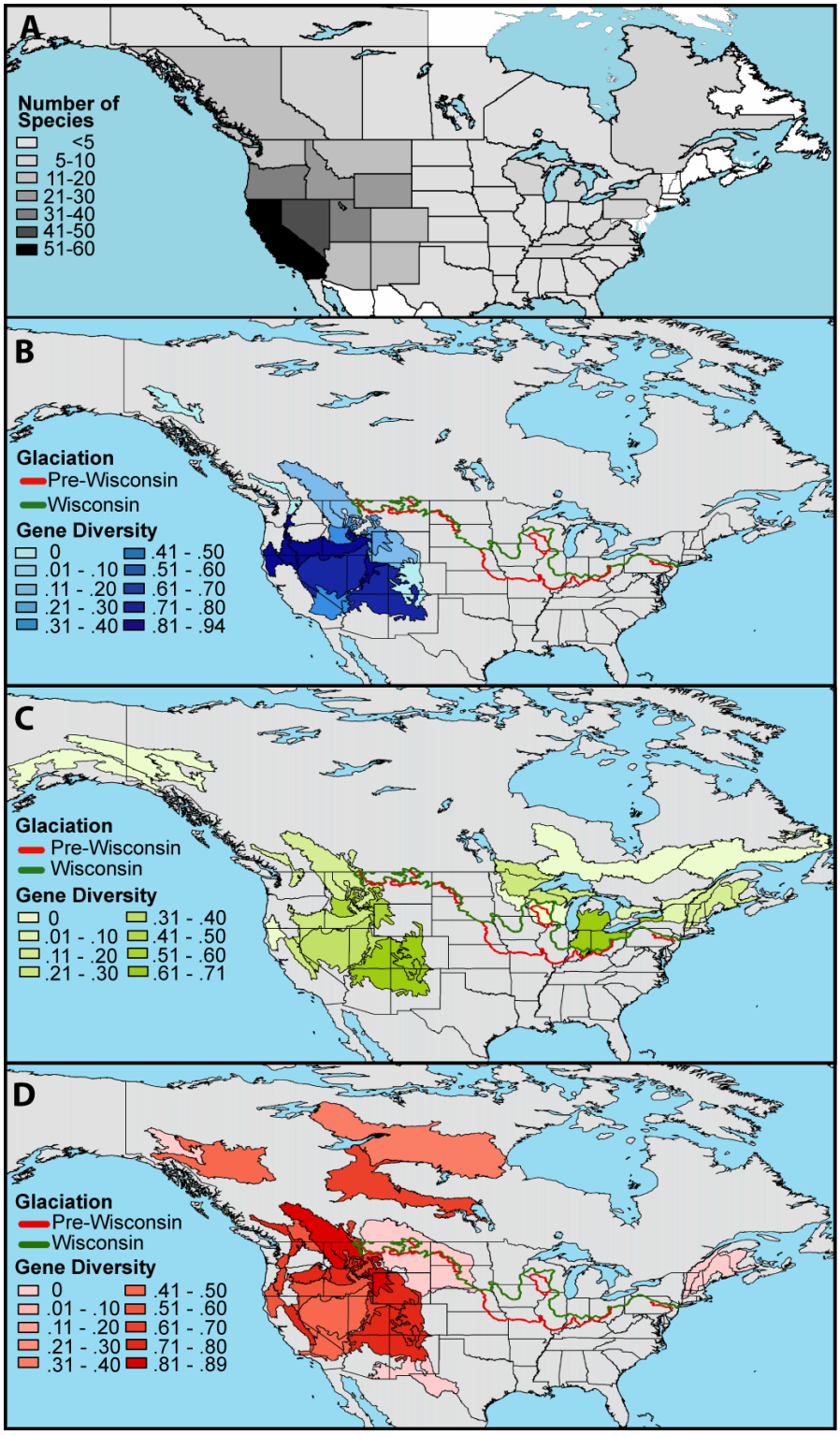

Figure 1 (A) Species diversity calculated as number of species per state (B) Gene diversity lineage I (C) Gene diversity lineage II (D) Gene diversity lineage III; increasing colour depth = increasing gene diversity. Evolutionary lineages are shown and correspond to Fig. 2.

DEFINITION OF HAPLOTYPES, PSEUDOGENES AND SUPRAHAPLOTYPES

The analysis of the *trn*L-*trn*F (total alignment length 1148 bp, see supplement Table 2 for the alignment) plastidic region allowed us to identify 241 cpDNA haplotypes.168 of those were found to be singletons and 92 haplotypes were shared by several accessions. Haplotype coding was done according to Dobeš et al. (2004a) (haplotypes A to DU) and Schranz et al. (2005) (haplotypes DV to GS) (see supplement Table 3 for haplotype names and genebank accession numbers). Up to 3 *trn*F pseudogenes in the *trn*L-F intergenic spacer were found adjacent to the functional *trn*F gene between alignment positions 785 and 1100 (copy I 785-900, copy II 901-991, copy III 992-1000). Haplotypes differing only in pseudogene number or mutations in this region were collapsed into suprahaplotypes (supplement Table 4).

For a higher resolution and statistical support in the phylogenetic reconstruction and network analysis, 303 accessions from the *trn*LF dataset representing 127 haplotypes were combined with *rpo*C1 sequences. The total length of the *rpo*C1 alignment alone was 821 bp, the combined alignment had a total length of 1972 bp (alignment see supplement Table 5). The additional variable characters in the *rpo*C1 region resulted in the definition of 147 combined haplotypes. Combined *trn*LF/*rpo*C1 haplotypes were given numbers which are listed in supplement Table 6 together with the corresponding *trn*LF haplotype name and the genebank accession numbers of the *rpo*C1 types.

LOW RESOLUTION IN THE BACKBONE: *RPO*C1 AND *TRN*LF TREES

Resolution of the phylogenetic trees obtained from the analyses of *trn*LF (Figure 4, online material) was low. However it could be enhanced by the combination of the *trn*LF and *rpo*C1 dataset (Figure 5, online material). Both new and previously described evolutionary lineages (I, II and III as in (7)) could be identified, and grouping between them was used to divide the dataset for subsequent network analyses.

NETWORK STRUCTURE REVEALS SIX ANCESTRAL cpDNA LINEAGES

Network analysis of the *trn*LF dataset revealed a complex pattern of lineages as well as several starlike nodes in the network (Figure 2).

Suprahaplotype 8 was found to be in the centre of the network to which all identified lineages (I to VI) were connected. Lineage I consisted of two frequent suprahaplotypes (S1 and S7) and a number of haplotypes and suprahaplotypes derived from them. The same was true for lineage II, although the network structure of this lineage was considerably less complex than that of lineage I. Most accessions

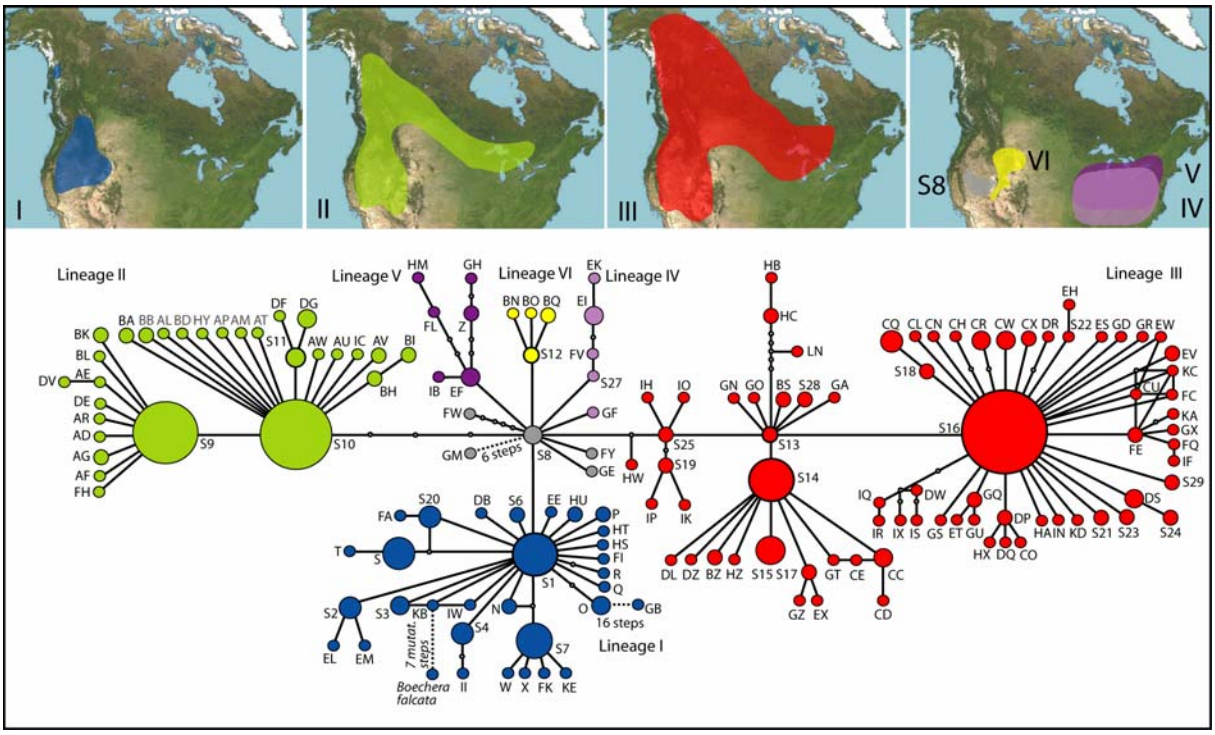

Figure 2 Network Analysis based on *trn*LF chloroplast DNA; the network was reconstructed using TCS1.21. Node size corresponds to number of accessions carrying the (supra)haplotype. Portions of the network are coloured according to their lineage identity. The same colour code was used for displaying the distribution ranges in the top figures.

in lineage II were found either in S9 or S10. From these two suprahaplotypes a range of haplotypes differentiated by a single mutation step were derived, a pattern consistent with rapid range expansion (Koch et al. 2006). In Lineage III the frequency of suprahaplotypes increased with the distance from the centre of the network, an unusual finding as central/old haplotypes are typically most frequent. The most frequent haplotypes were all derived from each other. From S13, S14 and S16 several haplotypes were derived, giving the same pattern as found in lineage II for S9 and S10. The network analysis of the combined dataset for *trn*LF and *rpo*C1 yielded the same result, although the addition of the *rpo*C1 sequence produced additional mutation steps due to more variable characters (Figure 6, online material).

HAPLOTYPE LINEAGE DISTRIBUTION RANGES BROADLY OVERLAP

The distribution ranges of the cpDNA lineages I, II, III and VI overlapped in western North America (the Great Basin and adjacent mountain ranges). Lineage I occurred in the main distribution range of *Boechera* and only 15 individuals carrying lineage I haplotypes were found in the Yukon and the Northern Rocky Mountains. Lineage II (suprahaplotypes 9 and 10) and III (suprahaplotypes 13, 14, 15) were also found in the range of lineage I, but also occurred north of the last glaciation. The most derived haplotypes within these lineages were nonetheless found in more southern regions. Lineage VI was distributed in the north-eastern range of lineage I (Figure 2).

PATTERNS OF GENETIC DIVERSITY ARE DIFFERENT IN LINEAGES I, II AND III

Centres of gene diversity were found to be different for lineages I, II and III, and followed a slight west to east trend. Lineage I had its centre of gene diversity in the Klamath-Siskiyou region north of the Sierra Nevada (p=0.94), followed by the the Snake/Columbia Plateau and the Wasatch/Uinta Range (p=0.89 and p=0.85) (figure 1b). The centre of gene diversity for lineage II was was found to be the Colorado Plateau and the adjacent Colorado Rocky Mountains (p=0.78), and decreased towards the west, north and south (Figure 1c).

For lineage III the highest gene diversity was found in the southeastern central Rocky Mountains (Figure 1d) (p=3.0), and followed the Sierra Nevada and the other mountains ranges surrounding the Great Basin.

In general the gene diversity was lowest in regions north of the last glaciation. The highest number of singletons in lineage I and II was found in the centre of gene diversity, while in Lineage III this was found in the Sierra Nevada which had the second highest gene diversity.

RANGE EXPANSION, STABLE POPULATIONS AND MULTIPLE COLONIZATIONS: PHYLOGEOGRAPHIC ANALYSIS BY THE COMPARISON OF GENE AND NUCLEOTIDE DIVERSITY

Following Avise (2000) the comparison of gene (h) and nucleotide (p) diversity allows the inference of possible phylogeographic scenarios, including: (A) high h and p – stable population over long term with large Ne (effective population size) or multiple colonizations and admixed populations, (B) high h and low p – rapid population growth from small ancestral population with small Ne, (C) low h and high p – transient bottleneck in large ancestral population or admixture from small, geographical subdivided population and (D) low h and low p – rapid long distance dispersal and migration or prolonged or severe bottleneck. For regions south of the last glaciation the derived phylogeographic scenario often differed for lineages I, II and III within one region (Figure 3, and supplement Tables 7 and 8). For regions north of the last glaciation we inferred rapid range expansion except for the ecoregion south of the Great Lakes for lineage II.

SPECIES AND HAPLOTYPE-DIFFERENTIATION – TWO UNLINKED PROCESSES?

Our network analysis demonstrated a mixture of few species specific lineages and suprahaplotypes that are shared by up to 27 taxa. True species specific lineages were rare. In lineage I suprahaplotype 20 and its derivatives represented a lineage specific to *B. microphylla*. Lineage II was dominated by *B. stricta* individuals. Lineage IV was specific for *B. canadensis*, and lineage V was characterized by several Eastern North American *Boechera*, most of them with private haplotypes (*B. laevigata*, *B. missouriensis*, *B. perstellata*, *B. serotina*). Lineage III contained

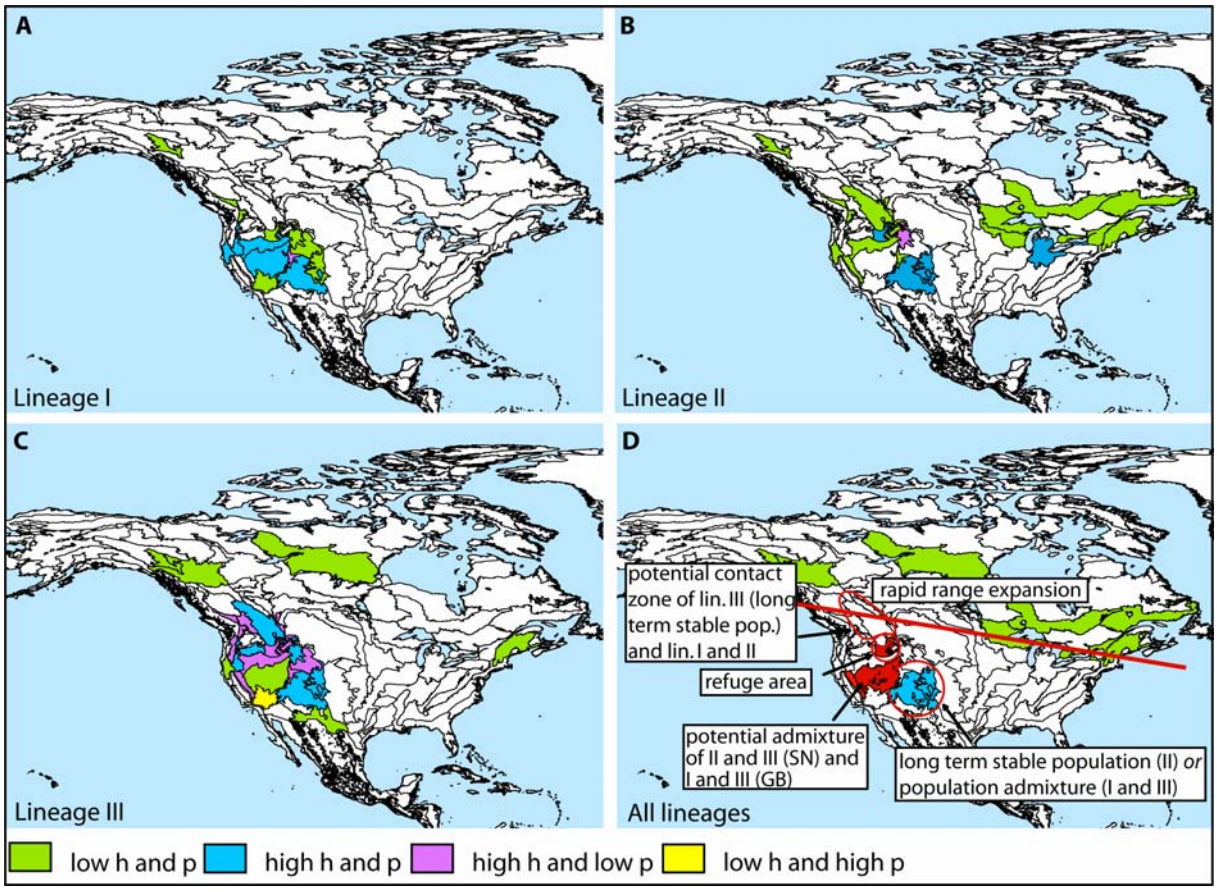

Figure 3 Phylogeographic scenarios derived according to Avise (2000). A-C Scenarios inferred for lineages I, II, III, D Summary of the most important phylogeographic implications.

only one species-specific lineage ((*Boechera fendleri* var. *spatifolia*, now *Boechera spatifolia* (14)). Suprahaplotypes 2, 9, 10, 11, 19 and 20 were shown to be species-specific or predominantly found in one species.

Discussion

Phylogeographic studies often consider only one single species because different species may have different habitat preferences and life cycles and hence may show a different response to climatic events such as the Quaternary ice ages. If several species are considered phylogeographic histories of species are first inferred for single species and later compared to each other (comparative phylogeography). For *Boechera* we tested if it is possible to conduct a phylogeographic study simultaneously for 57 taxa. Our network analysis based on cpDNA markers *trnL-F* and rpoC1 revealed six evolutionary lineages all arising from a central suprahaplotype. Comparing haplotype identity and species identity showed that haplotype sharing is very common within *Boechera*. This can either be explained by frequent hybridisation which was previously reported for *Boechera* or by the fact that haplotype differentiation predates speciation. Both possibilities seem to hold true for

Boechera. Since the cpDNA haplotypes are older than the species in many cases they represent an old gene pool. Therefore we conclude that a phylogeographic study based on this cpDNA marker system can be conducted for the whole group of 57 taxa simultaneously if the dataset is subdivided according to the evolutionary lineages inferred.

Phylogeographic analysis was done by the comparison of gene and nucleotide diversity (Avise, 2000) for the major evolutionary lineages I, II and III separately. Lineages IV and V were treated separately since they comprise only species restricted to Eastern North America (Kiefer et al. submitted). Gene diversity is based on the number of different DNA sequences and their frequency in a region while nucleotide diversity is a measure of differences among haplotypes within one region. The higher the gene diversity, the more different sequences were detected, the higher the nucleotide diversity, the more distantly related are the haplotypes. If gene and nucleotide diversity are both high this is taken as an indication for a long term stable population because new haplotypes had time to evolve (high gene diversity) and differentiate (high nucleotide diversity). Another possible interpretation may be multiple colonization events into this region also because several (high gene diversity) haplotypes which had previously differentiated (high nucleotide diversity) arrived from different source regions. On the other hand a region where gene and nucleotide diversity are low has either experienced a severe bottleneck during which most of the genetic diversity was lost or recent rapid range expansion where only a limited number of haplotypes reached the colonized area and time was too short for differentiation.
Our analysis revealed that the phylogeographic history of the evolutionary chloroplast DNA lineages I, II and III is different although their distribution ranges overlap in most parts. The lineages differed in respect to the centres of gene diversity and also in their distribution of singleton haplotypes. However, all three lineages had their lowest gene and nucleotide diversity north of the last glaciation indication recent range expansion. This is a common pattern found in phylogeographic studies often referred to as southern richness versus northern purity caused by the retreat of taxa into refugia south of the ice shield during the quaternary ice ages (Hewitt, 2001). One exception was the Great Lakes region where we detected a high gene diversity for lineage II. This was already noted in the previous phylogeographic study and interpreted as a potential glacial refuge in this region (Dobeš et al. 2004a). The eastern Great Lakes region was also shown to be a glacial refuge for butterflies (*Lycaeides*; Nice et al. 2005).A small isolated range of lineage I haplotypes in the Yukon interior dry forest in western Canada should also be noted. Only individuals carrying haplotype M included in suprahaplotype 1 (S1) were found in this northern enclave separated from the distribution range of lineage I by several hundred kilometres. Haplotype M is an old haplotype since S1 is central to lineage I. Its

presence in the Yukon interior dry forest might indicate a larger preglacial distribution range of lineage I reaching into the north of Canada. However, long-distance dispersal of this haplotype from the Cascades or the central Rocky Mountains where it is most abundant cannot be ruled out.

South of the last glacation no common pattern was found for the three evolutionary lineages. Klamath-Siskiyou was found to be the centre of genetic diversity for lineage I. This ecoregion is known for its extraordinary biodiversity and it was not or at least not heavily affected by the quaternary ice-ages (Ricketts et al. 1999). Therefore it may have offered a suitable habitat to *Boechera* over a long period of time. Comparison of h and p indicates a long term persistent population for lineage I in Klamath/Siskiyou, which also hosted the largest number of singleton haplotypes. However, for lineages II and III rapid long distance dispersal was inferred, thus illustrating the different phylogeographic histories of each evolutionary lineage. Considering the haplotype distribution ranges it is likely that lineage III invaded Klamath/Siskiyou from the Sierra Nevada (shared haplotype FE) or the central and southern Cascades postglacially. The same may hold true for lineage II. This leads us to the conclusion that Klamath/Siskiyou possibly offered possibilities for hybridisation between lineage I and lineage II or III haplotype carriers.

The Colorado Plateau was the centre of genetic diversity for lineage II and both gene and nucleotide diversity were high. Evidence for a long term stable population of lineage II also comes from the presence of singleton haplotypes not only on the Colorado Plateau but especially in the adjacent Colorado Rocky Mountains. The Colorado Rocky Mountains were described as a refuge for limber pine (Mitton et al. 2000) which fits well with a long term persistent population of *Boechera* lineage II. On the other hand the Colorado Rockies were also described as a genetically poor area (desert spider, Ayoub and Riechert 2004), a scenario which also applies to lineage I which was genetically poor in the Colorado Rocky Mountains. A high gene diversity was found for lineages I and III in the Colorado Plateau, although a high gene and haplotype diversity can be interpreted in two ways; a long term stable population or multiple colonization events. The latter interpretation seems to be true for lineages I and III because the suprahaplotypes present in the region do not form a continuous lineage but have missing “connecting” suprahaplotypes between them. The long term persistence of lineage II and the possibility of several subsequent colonization events by lineage I or II haplotype carriers offers multiple possibilities of hybridisation between individuals carrying haplotypes from different lineages.

The southern central Rocky Mountains comprise a region which has been addressed numerous times in phylogeographic studies in the past. In our analysis the southern central Rocky Mountains were divided into an eastern and a western section. The southeastern Rocky Mountains were shown to be the centre of genetic diversity for lineage III. Gene diversity and nucleotide diversity indicated a long term stable population or multiple colonization events. The eastern part of the southern central

Rocky Mountains has only been described as a glacial refuge for limber pine (Mitton et al. 2000) which might be explained by the fact that in the Eastern part of the southern central Rocky Mountains glaciation is assumed to have started in the Windriver Range (Leonard 2007). The high diversity in this ecoregion seems to be influenced by haplotypes that are also found in more southerly regions of the Rocky Mountains like the Colorado Rockies or the Wasatch/Uinta Mountains.

The southwestern central Rocky Mountains are separated from the southeastern central Rocky Mountains by the Montana Valley and Foothill Grasslands which are in parts characterized by a semi-dry landscape similar to that found on the Snake/Columbia plateau (Olson et al. 2001). The gene diversity in *Boechera* was lower in the southwestern central Rockies than in the southeastern central Rockies and the percentage of singleton haplotypes was about the same. However, nucleotide diversity was lower, so we rapid population growth from ancestral population with small Ne (Avise 2000) for this ecoregion. This may indicate a glacial refuge for *Boechera* with a small remaining population from which it expanded rapidly after deglaciation within this part of the Rocky Mountains chain. This hypothesis is likely as the western part of the southern central Rocky Mountains (Salmon River Mountains, Clearwater Range) is known as a refuge area (e.g. *Salix melanopsis* (Brunsfeld et al. 2007); *Cardamine constancei* (Brunsfeld and Sullivan 2005); red-tailed chipmunk (Good and Sullivan 2001); pine beetle (Maroja et al. 2007)) and as genetic contact zone (Brunsfeld and Sullivan 2005).

In terms of haplotype composition the southwestern central Rocky Mountains are similar to the northern central Rocky Mountains. The northern central Rocky Mountains stick out among other ecoregions as it is the only one in which all major suprahaplotypes from lineage III are present, thus making it a potential centre of origin or refuge area. However, only one singleton haplotype occurs here compared to the Sierra Nevada, where nine singleton haplotypes occur in addition to all but one suprahaplotype. Therefore we hypothesize that lineage III evolved in the Sierra Nevada, and from there colonized today's distribution range to reach the northern central Rocky Mountains via the southwestern central Rocky Mountains. It may have survived in the Bitterroot Range or adjacent ranges, where S15 (the suprahaplotype absent from the Sierra Nevada) evolved to become the major postglacial colonization haplotype.

In summary our continent-wide phylogeographic study of *Boechera* shows that within this genus there are three lineages with a unique phylogeographic history which shows migration and colonization for a time predating speciation in this genus. Migration seems not to have happened only one way. As differentiation in a lineage proceeded haplotypes moved further into new habitats and at the same time backwards into their old range leading to the overlap of distribution ranges of haplotypes today.

In comparison to other phylogeographic studies the results for *Boechera* are in agreement with some of them while they are in conflict with others. *Boechera*'s phylogeographic history is in concordance with patterns which were often revealed in the past such as the glacial refugium in the southern central Rocky Mountains. More phylogeographic studies covering a larger area would be needed for a better comparison or for analysing common patterns as it was done for eastern North America (Soltis et al. 2007).

In order to connect the obtained gene diversity data to their biological background we calculated species diversity per state and compared haplotype and species identity. California was identified as the state with the highest species diversity for *Boechera,* even when apomictic species were excluded from the analysis. California is known for its botanical biodiversity and is among the 25 biodiversity hotspots on Earth (Calsbeek et al. 2003). The high biodiversity in California can be explained by the presence of multiple habitat types (elevation belts of the Sierra Nevada) which may also be the reason for the high species diversity in *Boechera*. The largest portion of the adjacent state of Nevada, which is second in *Boechera* biodiversity, is covered by the Great Basin, a highly structured landscape which displays a mixture of cold-temperate and desert vegetation which shows the influence of the adjacent Mojave Desert (Ricketts et al. 1999). This particular mixing of species with different climate preferences was also true for *Boechera* (see supplement table 9) and is correlated with the high species diversity in Nevada.

Comparison of species and haplotype identity revealed few species specific haplotypes or lineages, in conjunction with elevated levels of haplotype sharing. Suprahaplotype 8, which is central to the network and connects all lineages is the oldest suprahaplotype according to coalescent theory (Posada and Crandall 2000). It occurs only in the southern Great Basin and is shared by several species, and has been estimated to be 0.7 to 2 million years old (Dobeš et al. 2004a). This places the origin of all *Boechera* lineages and also all derived haplotypes into the pleistocene. While age estimates for other haplotypes are imprecise and characterized by large confidence intervals, the relative ages between haplotypes can nonetheless be compared. Species-specific lineages connected directly to S8 are likely relatively old species as they represent the earliest split possible within *Boechera*. These early diverged species-specific lineages include lineage II (specific to *B. stricta*; haplotypes are rarely shared), lineage IV (specific to *B. canadensis*), and lineage V which represents another eastern North American branch within *Boechera* (*B. laevigata*, *B. missouriensis* and *B. shortii*). (Supra)haplotypes specific to *Boechera microphylla* and *Boechera spatifolia* were derived from suprahaplotypes further away from the centre of the network, thus implying that these species have a more recent origin.

Although some species- and lineage-specific haplotypes were detected, they are the exception since haplotype sharing is the rule. One explanation for haplotype sharing could be frequent hybridisation and subsequent backcrossing, which may lead to chloroplast capture. Alternatively the age of some species may be so young that significant differentiation between haplotypes has not yet happened. This seems to be well illustrated by S16 (lineage III) which is shared by 50% of the taxa in our analysis.

Tracing species abundance through the suprahaplotypes in the network, it is remarkable that one group of species appears in all major nodes except for the species-specific lineages. Those species are - according to Rollins' taxonomic system (Rollins 1993) – the five varieties of *B. holboellii* which have all obtained species rank (Windham and Al-Shehbaz 2006). Their presence in all major nodes of the network means that these species or former varieties encompass most of the chloroplast DNA genetic variability found in the genus *Boechera*. We therefore think it is likely that *Boechera holboellii* sensu Rollins represents the ancestral types from which most other *Boechera* species evolved.

Plotting apomictic species on the network revealed that they are not found within one lineage but rather scattered across the tree. There were indications from microsatellite studies that apomixis arose repeatedly from a sexual background after hybridisation (Sharbel et al. 2001; Dobeš et al. 2004b). At least the repeated independent expression of apomixis is also supported by chloroplast DNA data. However, since chloroplasts are maternally inherited it may be more precise to say that there is no single maternal apomictic lineage. Since chloroplast differentiation predates speciation in most cases in *Boechera* nuclear data would be needed to really prove the absence of apomictic lineages. This is work currently being in progress.

Summary and Outlook

Based on cpDNA haplotype variation and differentiation we were able to reconstruct major processes in *Boechera* evolutionary history and biogeography. For some taxa these reconstructions were even possible on species-level. However, more research is needed to link species definition (alpha taxonomy) with phylogeny and biogeography. Our ongoing research and preliminary results based on ITS (internal transcribed spacer of nuclear ribosomal DNA) and various single copy genes indicate that simple gene trees and even coalescent theory might not help to resolve any species. It is more likely that a continental wide population genetics approaches will lead to a deeper understanding of the relationship of single *Boechera* species.

ACKNOWLEDGEMENTS

Funding over the last years from DFG (German Research Foundation) and ASF (Austrian Science Foundation) to M. A. K. was greatly acknowledged. We are very

grateful for suggestions and comments made by Dr. Ihsan Al-Shehbaz (Missouri Botanical Gardens/MO/USA).

LITERATURE CITED

Al-Shehbaz, I.A., submitted. *Boechera* A. Löwe and D. Löwe. In: Flora North America.

Avise, J., 2000. In: *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Reviews in Ecology and Systematics 18, 489-522.

Avise, J.C., 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7, 371-379.

Ayoub, N.A., Riechert, S., 2004. Molecular evidence for Pleistocene glacial cycles driving diversification of a North American desert spider, *Agelenopsis aperta*. Molecular Ecology 13, 3453-3465.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane Jr., S.L., Warwick, S.I., Windham, M.D., Al-Shehbaz, I.A.., 2006. Toward a global phylogeny of the Brassicaceae. Molecular Biology and Evolution 23, 2142-2160.

Brubaker, L.B., Anderson, P.M., Edwards, M.E., Lozhkin, A.V., 2005. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapping pollen data. Journal of Biogeography 32, 833-848.

Brunelle, A., Whitlock, C., 2003. Postglacial fire, vegetation, and climate history in the Clearwater Range, Northern Idaho, USA. Quaternary Research 60, 307-318.

Brunsfeld, S.J., Miller, T.R., Carstens, B.C., 2007. Insights into the Biogeography of the Pacific Northwest North America: Evidence from the Phylogeography of *Salix melanopsis*. Systematic Botany 32, 129-139.

Brunsfeld, S.J., Sullivan, J., 2005. A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conservation Genetics 6, 895-904.

Calsbeek, R., Thompson, J.N., Richardson, J.E., 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Molecular Ecology 12, 1021-1029.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9, 1657-1660.

Dobeš, C.H., Mitchell-Olds, T., Koch, M.A., 2004a. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* X *divaricarpa*, and *A. holboellii* (Brassicaceae). Molecular Ecology 13, 349-370.

Dobeš, C.H., Koch, M.A., Mitchell-Olds, T., 2004b. Intraspecific diversification in North American *Boechera* stricta (= *Arabis drummondii*), *Boechera* x divaricarpa, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. American Journal of Botany 91, 2087-2101.

Dobeš, C., Koch M., Sharbel, T., 2006. Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Annals of the Missouri Botanical Gardens 93, 517-533.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): An integrated software for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47-50.

Good, J.M., Sullivan, J., 2001. Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. Molecular Ecology 10, 2683-2695.

Hewitt, G., 2000. The genetic legacy of the Quaternary Ice Ages. Nature 405: 907-913.

Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Molecular Ecology 10, 537-549.

Hull, J.M., Girman, D.J., 2005. Effects of Holocene climate change on the historical demography of migrating sharp-shinned hawks (*Accipiter striatus velox*) in North America. Molecular Ecology 14, 159-170.

Jakob, S.S., Blattner, F.R., 2006. A Chloroplast Genealogy of Hordeum (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequences for Phylogenetic Inference. Molecular Biology and Evolution 23, 1602-1612.

James, L.A., Harbor, J., Fabel, D., Dahms, D., Elmore, D., 2002. Late Pleistocene Glaciations in the Northwestern Sierra Nevada, California. Quaternary Research 57, 409-419.

Jordon-Thaden, I., Koch, M. accepted. Diversity patterns in the genus *Draba*: A first global perspective. Plant Ecology and Diversity.

Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., de Vries, S., de Jong, H., 2007. Diploid apomicts of the *Boechera holboellii* complex show large scale chromosome substitutions and different aberrant chromosomes. Proceedings of the National Academy of Sciences 104, 14026-14031.

Koch, M., Dobeš, C., Mitchell-Olds, T., 2003.: Multiple hybrid formation in natural populations: Concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molecular Biology and Evolution 20, 338-350.

Koch, M., Kiefer, C., Vogel, J., Ehrich, D., Brochmann, C., Mummenhoff, K., 2006. Three times out of Asia Minor - the phylogeography of *Arabis alpina* L. (Brassicaceae). Molecular Ecology 15, 825-839.

Koch, M.A, Dobeš, C., Kiefer, C., Schmickl, R., Klimes, L., Lysak, M.A., 2007a. SuperNetwork identifies multiple events of plastid *trn*F (GAA) pseudogene evolution in the Brassicaceae. Molecular Biology and Evolution 24, 63-73.

Koch, M.A., Matschinger, M., 2007b. Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences 104, 6272-6277.

Leonard, E., 2007. Modelled patterns of Late Pleistocene glacier inception and growth in the Southern and Central Rocky Mountains, USA: sensitivity to climate change and paleoclimatic implications. Quaternary Science Reviews 26, 2152-2166.

Maroja, L.S., Bogdanowicz, S.M., Wallin, K.F., Raffa, K.F., Harrison, R.G., 2007. Phylogeography of spruce beetles (*Dendroctonus rufipennis* Kirby) (Curculionidae: Scolytinae) in North America. Molecular Ecology 16, 2560-2573.

Mitton, J.B., Kreiser, B.R., Latta, R.G., 2000. Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. Molecular Ecology 9, 91-97.

Nice, C.C., Anthony, N., Gelembiuk, G., Ratermans, D., Ffrench-Constant, R., 2005. The history and geography of diversification within the butterfly genus *Lycaeides* in North America. Molecular Ecology 14, 1741-1754.

Olson, D.M., Dinerstein, E., Wikramanayke, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: New map of life on earth. Bioscience 51, 933-938.

Osborn, G., Bevis, K., 2001. Glaciation in the Great Basin of the Western United States. Quaternary Science Reviews 20, 1377-1410.

Phillips, F.M., Zreda, M.G., Smith, S.S., Elmore, D., Kubik, P.W., Sharma, P., 1990. Cosmogenic chlorine-36 chronology for glacial deposits at Bloody Canyon, eastern Sierra Nevada California. Science 248, 1529-1532.

Posada, D., Crandall, K.A., 2001. Intraspecific phylogenetics: Trees grafting into networks. Trends in Ecology and Evolution 16, 37-45.

Ricketts, T.H., Loucks, C.J., Dinerstein, E., 1999. Terrestrial Ecoregions of North America: A conservation Assessment. Island Press.

Rollins, R.C., 1993. The Cruciferae of Continental North America. Stanford University Press, Stanford, CA.

Roy, B.A., 1995. The breeding system of six species of *Arabis* (Brassicaceae). American Journal of Botany 82: 869–877.

Schmickl, R., Kiefer, C., Dobeš, C., Koch, M.A., 2008. Evolution of trnF(GAA) pseudogenes in cruciferous plants. Plant Systematics and Evolution: DOI 10.1007/s00606-008-0030-2.

Schranz, M.E., Dobeš, C.H., Koch, M.A., Mitchell-Olds, T., 2005. Sexual reproduction, hybridization, apomixis and polyploidization in the genus *Boechera* (Brassicaceae). American Journal of Botany 92, 1797-1810.

Sharbel, T.F., Mitchell-Olds, T., 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (*Brassicaceae*). Heredity 87, 59-68.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15, 4261-4293.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, R.S., Anderson, K., 2000. Biomes of western North America at 18,000, 6,000 and 0 14C yr B.P. reconstructed from pollen and packrat midden data. J ournal of Biogeography 27, 555-584.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7, 453-464.

Voigt, M.L., Melzer, M., Rutten, T., Mitchell-Olds, T., Sharbel, T.F. In: Hörandl, E., Grossniklaus, U., Van Dijk, P., Sharbel, T.F., 2007. in Apomixis: Evolution, Mechanisms and Perspectives, eds Hörandl E, Grossniklaus U, Van Dijk P, Sharbel TF (Koeltz, Koenigstein, Germany), pp 235–258.

Windham, M.D., Al-Shehbaz, I.A., 2006. New and noteworthy species of *Boechera* (Brassicaceae) I: sexual diploids. Harvard Papers in Botany 11, 61-68.

Windham, M.D., Al-Shehbaz, I.A., 2007a. New and Noteworthy Species of *Boechera* (Brassicaceae) II: Apomictic Hybrids. Harvard Papers in Botany 11, 257-274.

Windham, M.D., Al-Shehbaz, I.A., 2007b. New and Noteworthy Species of *Boechera* (Brassicaceae) III: Additional Diploids and Apomictic Hybrids. Harvard Papers in Botany 12, 251-274.

Legends supplementary material

Figure 4 (online material) Parsimony analysis of the *trn*LF dataset calculated from 1000 shortest trees.

Figure 5 (online material) Parsimony analysis of the combined *trn*LF-*rpo*C1 dataset calculated from 1000 shortest trees.

Figure 6 (online material) Network analysis based on the combined *trn*LF-*rpo*C1 dataset. Lineages identified were the same as in the analysis based on *trn*LF only. The number of the combined *trn*LF-*rpo*C1 haplotype is given in the corresponding node. Accessions included and their *trn*LF haplotype names are given in supplement table 5.

Figure 7 (online material) Distribution of all samples included in this analysis.

Table 1 (supplement material) Accession List stating accession number, origin of herbarium voucher, geographic coordinates, haplotype evolutionary lineage and ecoregion in which the sample was placed.

Table 2 (supplement material) Annotated alignment of all *trn*LF haplotypes including pseudogene region.

Table 3 (supplement material) Genebank accession details of haplotypes and voucher information of the individual in which the haplotype was found first.

Table 4 (supplement material) Suprahaplotype codes and haplotypes included within them.

Table 5 (supplement material) Combined alignment of *rpo*C1 and *trn*LF

Table 6 (supplement material) Combined *trn*LF-*rpo*C1 haplotypes and accession numbers of individuals in which they were found

Table 7 (supplement material) Haplotype abundance and haplotype names for each ecoregion investigated and gene and nucleotide diversity based on those numbers; a separate table for every lineage identified is included. A column also states the phylogeographic scenario inferred according to (16).

Table 8 (supplement material) Phylogeographic scenarios inferred for ecoregions and lineages along with information on ecoregions and a comparison to other phylogeographic studies.

Table 9 (supplement material) Species Distribution in the United States and Canada. 1 indicates the presence of a species in a state or province. The colour code is given below the table.

2.5 ITS and Single Copy Gene Analyses

This manuscript will be submitted to American Journal of Botany.

Supplementary material is given on the DVD in the back of the thesis.

Analyses which are in progress for the manuscript submission include Bayesian analyses for all datasets used in this study and a lineage through time plot for the ITS dataset.

Running Title:

ITS and single copy gene intron studies in *Boechera*

Molecular systematics and evolutionary history of North American *Boechera* (Brassicaceae) – nuclear encoded ITS and single copy non coding intron sequences characterize old lineages and single species.

CHRISTIANE KIEFER, CHRISTOPH DOBEŠ*, AND MARCUS A. KOCH#

Heidelberg Institute of Plant Science, Biodiversity and Plant Systematics, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany

* Present address: Department of Pharmacognosy, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

 # Corresponding author:
Marcus A. Koch
Heidelberg Institute of Plant Sciences, Department of Biodiversity and Plant Systematics, University of Heidelberg, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany.
Phone: +49-6221-54-4655
FAX: +49-6221-54-5508
eMAIL: mkoch@hip.uni-heidelberg.de

Manuscript received _____; revision accepted _____.

Acknowledgements
We are grateful to the curators of the herbaria GH, MO and DAO for providing leaf material, and Ihsan Al-Shehbaz for continuous advice and support during our studies on *Boechera* .

Abstract:
110 species are currently described for the North American genus *Boechera* (Brassicaceae). Several hybrid combinations are assumed based on morphological evidence. However, little is known about the phylogenetic relatedness of the species to each other. We used nrDNA ITS and intron sequences of two single copy genes to obtain deeper insights into *Boechera* phylogeny and ITS type distribution across taxa. We could show that species specific lineages exist although the relationship among them is poorly resolved. Comparing tree topologies this indicates rapid speciation which probably happened in the second half of the quaternary. Hybrids could be identified by the comparison of the different marker systems together with chloroplast DNA types from an earlier study.
We hypothesize that a continental wide population biology approach would be needed to unravel the complex relationships of evolutionary lineages in this diverse genus.

keywords: *Boechera* / Brassicaceae / ITS / single copy genes / hybridisation

INTRODUCTION

As complex as the geography of the North American continent is the genus *Boechera* (Brassicaceae) living in this heterogeneous environment. *Boechera* is part of the tribe Boechereae that encompasses seven closely related genera, *Anelsonia, Cusickiella*, *Nevada*, *Phoenicaulis, Polyctenium*, *Sandbergia* (= *Halimolobus perplexa*, Al-Shehbaz, 2007) (Bailey et al. 2006). They are all almost exclusively North American and have a base chromosome number of x=7 (Dobeš et al., 2006, Warwick et al., 2006) and live in a wide range of habitats reaching from desert and shrub steppe to montane meadow and habitats above timberline.

.With few exceptions members of the tribe Boechereae are characterized by branched trichomes and entire leaves. Most of them are perennials with a well-developed basal rosette (Al-Shehbaz & al., 2006). The taxonomic history of *Boechera* has been eventful. For a long time the taxa were treated as members of the genus *Arabis*. *Arabis holboellii* was described by Hornemann in 1827 (Hornemann, 1827) and later transferred into the genus *Boechera* which was established in 1976 with *Boechera holboellii* as type species (Löve and Löve, 1976). However, when North American *Arabis* species were revised, such as *Arabis retrofracta*, Graham, they were described as varieties of *Boechera holboellii* which was then called *Arabis* again, in this case *A. holboellii* (Hornemann) A. Löve & D. Löve var. *retrofracta* (Graham) Rydberg. Most detailed descriptions of the species were available from Rollins (1993) who treated today's *Boechera* as a separate series of *Arabis* (Rollins, 1941). Analyses based on molecular data finally showed that *Arabis* was polyphyletic (Koch et al. 1999, 2000) and subsequently taxa related to *Arabis drummondii* (sensu Rollins, 1993) and *Arabis holboellii* (sensu Rollins, 1993) were transferred into *Boechera* (Al-Shehbaz, 2003). *Boechera holboellii* (sensu Rollins 1993) was a highly polymorphic taxon with a number of varieties with chromosome numbers of 2n=14 and 21 and a range of aneuploids (Rollins, 1941; Sharbel and Mitchell-Olds, 2001), and as in other taxa diploids and triploids were lumped into single species (Rollins, 1993). This view changed and in the forthcoming treatment of FNA (Flora North America) *B. holboellii* is restricted to Greenland (Al-Shehbaz, pers. comm.).

The formation of hybrids is a frequent phenomenon in *Boechera*, with homoploid hybridisation leading to the formation of diploid hybrids which often show lack of meiotic reduction of gametes and thus provide the first step in triploid formation (Dobeš et al. 2006, 2007a, 2007b). Imbalanced chromosome numbers in triploids preclude balanced meiosis and hence sexual reproduction, and thus triploid *Boechera* are apomictic (asexual reproduction via seed). Interestingly, diploid apomixis, which is virtually absent in the plant kingdom (e.g. in *Paspalum rufum*, Siena et al. 2008), is also found in *Boechera* (Böcher 1951).

Boechera was completely revised for the new flora of North America (Al-Shehbaz, unpublished; Windham and Al-Shehbaz, 2006, 2007a, 2007b), and in this most recent treatments of the genus apomictically and sexually reproducing taxa have been separated resulting into 72 species which reproduce sexually and 38 species characterized by apomixis.

In the past few years various attempts have been started to unravel also the evolutionary history and systematics of more widely distributed members of *Boechera* such as *B. stricta* (= *B. drummondii*) and *B. holboelllii* (Koch et al. 2003, Dobeš et al. 2004a, 2004b). These studies were extended including other taxa (Schranz et al. 2005) also employing a broader spectrum of molecular and cytogenetic approaches (Song et al. 2006).

Studies based on chloroplast DNA gave insights into past events which shaped haplotype distribution in *Boechera* such as post-glacial range expansion or past range fragmentation (Dobeš et al. 2004a, 2004b). Further analyses including more *Boechera* taxa were consistent with those findings and revealed a large amount of haplotype sharing among species which was attributed either to the haplotype differentiation predating speciation and incomplete lineage sorting and/or frequent hybridisation (Dobeš et al. 2004a, Kiefer et al. submitted).

Based on the previously obtained insights into the evolution of the maternal lineage (Dobeš et al. 2004a, Kiefer et al. submitted) we wanted to apply biparentally inherited markers in order to get an insight into speciation processes. In our analysis we employed nrDNA ITS1 and ITS2 as well as intron sequences from genes neighbouring the ELF3 locus (At2g25920; according to the ancestral crucifer karyotype from Schranz et al. 2007 this locus should be in genome block h which in *Boechera stricta* is on chromosome 5) and neighbouring the VRN1 locus (At3g18900, *Boechera stricta* chromosome 3; Schranz et al. 2007). VRN1 and ELF3 are both genes potentially involved in adaptation, so they may have an evolutionary significance. *VRN* genes mediate vernalization, the process by which a long period of cold induces a mitotically stable state that leads to accelerated flowering during later development. *VRN1* encodes a protein that binds DNA in vitro in a non-sequence-specific manner and functions in stable repression of the major target of the vernalization pathway (Levy et al. 2002). ELF genes are involved in plant circadian clocks (elf = early flowering) and are found with different variants (Hazen et al. 2005).ELF3 encodes a circadian clock-regulated nuclear protein that functions in an Arabidopsis PHYB signal transduction pathway (Liu et al. 2001).

We chose sequences from neighbouring genes because VRN1 and ELF3 are members of gene families while the adjacent genes were present as single copies only. That way we prevented the amplification of paralogues. In the following we refer to those genes by their locus (ELF3 neighbouring gene = At2g25920 and At3g18900 = VRN1 neighbouring gene) since their proteins have unknown function.

By choosing on the one hand nrDNA ITS1 and ITS2 and on the other hand single copy gene introns we obtained two entirely different marker systems. The ITS on the one hand is known to have the possibility to undergo concerted evolution (reviewed by Elder and Turner 1995). Concerted evolution means that over time the thousands of copies of ITS are adapted. Since the ITS is biparentally inherited this means that the ITS copies either adapt to the maternal type, to the paternal type or to a mixture of both types. However, also both copies may be conserved (Feliner et al., 2004). This characteristic of the ITS is of special interest in *Boechera* since a lot of hybrids can be expected that could carry either one of the parental ITS types, a recombinant ITS type or several ITS types. Indeed it was shown in the past that single *Boechera* individuals carried up to eight different ITS types (Koch et al. 2003).
Since the ITS is inherited biparentally it may be interpreted as a marker of present or past gene flow (Koch et al. 2005). Hence, biogeographic analyses of the ITS data may show which areas were in contact in the past or where gene flow persists to the present day. The pattern of ITS type distribution along with species identity may suggest if range fragmentation and expansion had an effect in speciation.
Integrative analysis of the different marker systems along with biogeographic analyses may help us understand how *Boechera* shaped its biodiversity.

MATERIALS AND METHODS

Plant material.—Leaf material was obtained from herbarium specimens from GH, MO and DAO. Corresponding accession details are listed in supplementary table 1, the geographic distribution is given in supplementary figure 1. In total we analysed 964 vouchers including outgroup specimens, but we did not obtain DNA sequence information for all accessions and all loci (refer to supplementary table 1).

DNA extraction.—Total DNA was obtained from a 0.5x0.5 cm^2 piece of dried leaf tissue from single individuals. Extraction followed the CTAB method of Doyle & Doyle (1987), but some modifications were applied, involving grinding of only a 0.5x0.5 cm^2 piece of dry leaf tissue in 2ml tubes using a Retsch swing mill (MM 200), addition of two units of ribonuclease (RNAse A) to the resolved DNA, and washing of the DNA pellet twice with 70% ethanol. DNA was finally dissolved in 50-70µl Bidest or low TE-buffer (Tris-EDTA) for long-term storage.

PCR conditions.—PCR reactions were performed in a volume of 25 µl containing 1x GoTaq buffer (Promega, Madison, USA), 2 mM $MgCl_2$, 5 pmol of each primer, 5 nmol dNTPs (1.25 nmol of each dNTP) and one unit Taq DNA polymerase (GoTaq, Promega), and variable concentrations of template (50 to 400 ng) using a PTC-200 thermal cycler (MJ-Research). Thermal cycling started with a denaturation step at

95°C lasting three min; followed by 30 cycles each comprising 30 s denaturation at 95°C, 30 s annealing at 48°C and elongation at 72°C. Amplification ended with an elongation phase at 72°C lasting 10 min, and a final hold at 4°C.
ITS1 and ITS2 were amplified as described in Dobeš & al. (2004b) with PCR products spanning the complete ITS1 and ITS2 as well as the intervening 5.8s rRNA gene. PCR products were checked on agarose gels (1% agarose in TAE).
As additional nuclear markers we chose intron sequences from genes flanking the VRN1 and ELF3 loci. Primer sequences for the intron in the single copy locus adjacent to *VRN1* were forward GCACTTGACCATCTCTTCAGATAA and reverse AGTCCTTCGACGCAAACTG. They are placed in the flanking coding regions and were obtained from Th. Mitchell-Olds and Eric Schranz, Duke University, USA. The intron in the gene neighbouring the *ELF3* gene was amplified using the primer sequences forward TTTGTTGTTGCATATGGTTGT and reverse TGCTTTACATGACTTGCTCTTA also optimized and obtained from Th. Mitchell-Olds and Eric Schranz, Duke University, USA. PCR products were checked on agarose gels (1% agarose in TAE). PCR products were all purified with the NucleoFastKit (Macherey-Nagel, Germany).

Cycle Sequencing.—Cycle sequencing was done with the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) using the PCR primers for the cycle sequencing reaction. Samples were resolved in 10 µl Loading Solution and then run on a MegaBace 500 Sequencer.

Alignments and ITS type definition.—Forward and reverse sequences were aligned, edited by hand and trimmed to a common length. New ITS-types were named following the nomenclature as introduced previously Koch et al. (2003) or were assigned to ITS-types already described and published by these authors (Koch et al. 2003; Dobeš et al. 2004b). An ITS type was defined as any sequencing differing from the other sequences by at least one mutation. Sequences with ambiguous sites were also named as an own ITS type since parents could not be determined in most cases by comparing to sequences without ambiguous sites and no cloning was done.
The alignments used for the phylogenetic analysis were made manually using the program GenDoc (Nicholas & Nicholas, 1997).
The alignment of the ITS sequences followed a previously published alignment (Koch et al. 2003). However, new gaps were introduced when needed. The ITS alignment was subdivided for the different types of analyses.
The alignments for the intron sequence were done by eye since sequence similarity was very high and no doubtful positions were present.

ITS, AT2G25920 and AT3G18900 phylogenetic tree reconstructions.— Phylogenetic reconstruction based on ITS was done twice with two different alignments. In the first analysis the alignment was once kept as a whole including all identified sequence types and also sequences with ambiguous sites. In the second analysis sequences with ambiguous sites and sequences originating from potential apomicts were excluded. We decided to follow this approach because ambiguous sites may lead to a false position of the accession in the tree. Also apomicts are assumed to be of hybrid origin and may therefore contain recombinant ITS types which would also alter the structure of the tree since they cause conflict in the data.

The analysis of both ITS alignments including ITS1 and ITS2 as well as the 5.8s rRNA gene was done in PAUP4.0beta (Swofford, 2002) running a parsimony analysis. The maximum number of trees retained was limited to 10,000. For the heuristic search sequences were added randomly in 1000 replicates during which 10 trees were saved in each replicate. Chuckscore was set to one. Gaps were treated as missing. TBR was used as branch-swapping algorithm. Starting trees were obtained via stepwise addition. All characters had equal weight.

For phylogenetic reconstructions based on the ITS dataset the alignment was supplemented with sequences of closely related genera obtained from Genebank as well as from our own dataset to determine the relationship of *Boechera* to other genera of the tribe Boechereae (AF146515, AF146514 = *Cusickiella douglasii*, DQ452059 = *Anelsonia eurycarpa*, DQ452061 = *Nevada holmgrenii*, DQ452066 = *Cusickiella quadricostata* and AJ232927, AJ232926 = *Halimolobus perplexa* var. *lemhiensis* (hereafter *Sandbergia*), AY230615 = *Polyctenium williamsiae*, and AF183109 = *Polyctenium fremontii*). *Capsella rubella* (AJ232913) served as appropriate outgroup.

Bootstrap analysis was also run in PAUP4.0beta with the same settings as the parsimony analysis in 1000 replicates.

Phylogenetic reconstruction based on the AT2G25920 intron data was done in the same way as with the ITS dataset with the exception that only *Polyctenium fremontii* and *Sandbergia perplexa* were used as representatives of other Boechereae. Polyctenium was used as outgroup.

For the AT3G18900 intron a Bayesian analysis was run using MrBayes through the program TOPALi (Milne et al. 2004). The SYM model was selected according to the model selection option in TOPALi, burn-in period was set to 25%, number of generations was set to 100,000 and number of runs was two. In this analysis *Cusickiella douglasii* and *Sandbergia perplexa* were included as additional species from the Boechereae. *Cusickiella* were used as outgroup.

In addition to the separate analyses a combined analysis including only accessions for which at least one single copy gene intron was sequenced was done using the same parsimony settings in PAUP 4.0beta as described above.

ITS network reconstructions.— Network reconstruction was done by running the program TCS1.21 (Clement & al. 2000) with the reduced alignment including only sequences without ambiguous sites or from individuals which were most likely apomicts according to the revision of the genus for the new Flora of North America (Al-Shehbaz, unpublished). Gaps were included as 5th state , and the connection limit was set to 95%. An initial analysis on this alignment resulted into a network with numerous circular connections. Therefore, the analysis was run on subsets of the alignment representing significantly supported subgroups recovered in the parsimony analysis.

Sequence analysis of subgroups.— In order to get an insight into the divergence of sequences within lineages, the original alignment was divided into subgroups according to the phylogenetic tree based on the ITS dataset. The analysis of the alignment was also done by using TOPALi with the display summary option (Milne et al. 2004).

Geographical analysis.—All information from the herbarium vouchers was entered into the database BioOffice (Biogis Consulting, Version 2.0.4). Missing geographical coordinates were added according to the descriptions on the herbarium vouchers. Using BioOffice haplotypes were plotted on North America maps included in the ArcView Package Version 8.

RESULTS

ITS-type definition.— In total 289 ITS types were detected among the 964 accessions and at least 63 species (only counting the ones which could without any doubt be assigned to a taxon defined by Al-Shehbaz, submitted) including ITS types characterized by ambiguous sites. Of those 39 were shared by several species, 30 were specific to one species and 220 were singletons. Of the 39 ITS-types shared by several species 21 were dominated by one taxon.

57 sequences contained ambiguous sites in forward and reverse sequence and were therefore assumed to be of hybrid origin; only three of them occurred twice, all others were singletons.

All genebank accession numbers, corresponding ITS-types, and accessions which share the ITS type and the ecoregion in which they were collected as well as cpDNA haplotype identity are shown in supplementary table 2. Colour codes indicate if the ITS type occurred once (yellow), if it was species specific (orange) or if it was shared by several taxa (red). Bold letters indicate the presence of ambiguous codes in the sequence.

Alignment analyses.— The ITS alignment including all sequence types detected in this study had a total length of 725 bp. 362 characters were constant, 154 variable characters parsimony-uninformative and 209 characters were parsimony informative (supplementary table 3). In total the alignment included 350 ITS sequence types representing *Boechera* accessions and 29 accessions representing the outgroup and other Boechereae.

The alignment which only included sequences without ambiguous sites and accessions which were most likely not apomictic (according to the taxonomic/cytological descriptions published by Windham and Al-Shehbaz, 2006, 2007a, 2007b; Dobeš et al. 2006) included only 168 ITS sequence types and 29 sequencing representing other Boechereae and the outgroup. The alignment had a total length of 721 base pairs including gaps (supplementary table 4). 439 of the characters were constant, and of the variable characters 143 were parsimony-uninformative and 139 were parsimony informative.

AT2G25920 was only sequenced for 210 accessions representing *Boechera* species and five accessions representing *Sandbergia* and *Polyctenium*. The total length of the AT2G25920 alignment was 564 bp. Six indels were coded separately and added as an AT-matrix to the alignment giving it a new total length of 570 bp (supplementary table 5). 331 characters were constant, 79 variable characters were parsimony-uninformative while 154 characters were parsimony-informative (with gap coding 160).

AT3G18900 was only sequenced for 111 accessions representing *Boechera* and four accessions representing *Sandbergia* and *Cusickiella*. The AT3G18900 alignment had a total length of 631 characters. Four gaps were coded in an AT matrix increasing the alignment length to 635 bp (supplementary table 6). 494 characters were found to be constant (including gap coding 495), 55 of the variable characters were parsimony-uninformative while 82 (including gap coding 86) were parsimony informative.

Phylogenetic reconstruction based on ITS.— The phylogenetic reconstruction based on ITS separated genera of the Boechereae as well as *Boechera* species into several major lineages (figure 1). However the relationship between the genera and within *Boechera* remained unresolved, apart from the bulk of *Cusickiella* accessions being sister to all other genera included in this study. Eastern North American *Boechera*, *Anelsonia*, *Nevada*, *Polyctenium*, *Sandbergia*, one *Cusickiella* accession and also *Boechera repanda* were represented by lineages arising from a polytomy together with a large of group of *Boechera* species centred in western North America. Within the large *Boechera* group 18 lineages containing more than one ITS-type arose from the polytomy. 44 ITS-types were not assigned to any lineage but were found separate on the polytomy. Some of the recovered lineages were species-specific while other lineages were dominated by one species and again others contained a multitude of species.

The most frequent ITS-types were generally shared by several species, though sometimes dominated by one or two taxa.

Taxa dominating in one group of ITS types or a single ITS-type were *Boechera puberula, Boechera subpinnatifida, Boechera sparsiflora, Boechera breweri, Boechera koehleri, Boechera stricta, Boechera pendulocarpa, Boechera lemmonii, Boechera retrofracta, Boechera microphylla, Boechera perennans, Boechera pendulina, Boechera demissa* or *oxylobula, Boechera cobrensis, Boechera platysperma, Boechera pulchra, Boechera glaucovalvula, Boechera rectissima, Boechera constancei* and *Boechera davidsonii*.

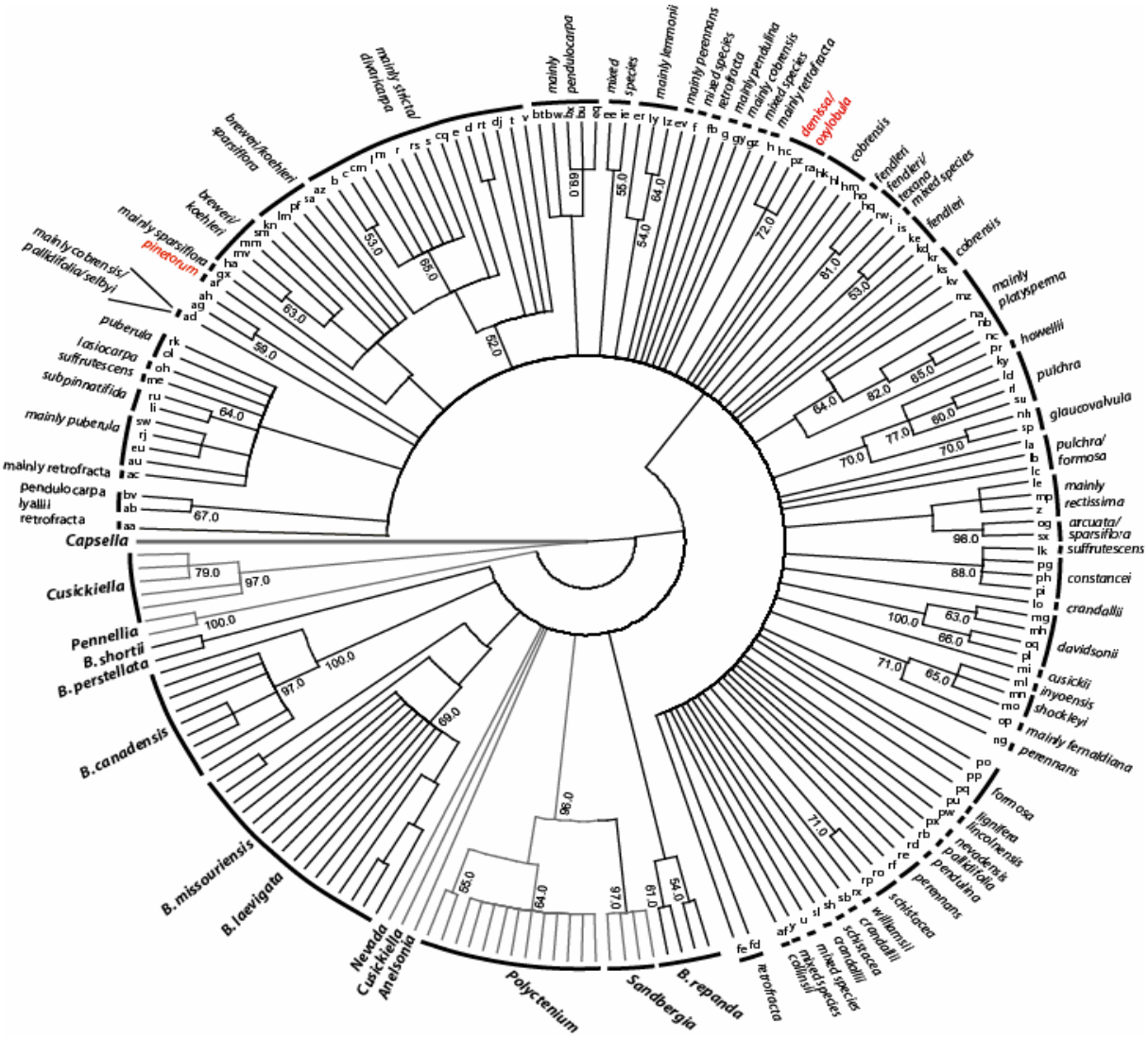

Fig.1 Parsimony analysis of the ITS dataset reduced to sequences without ambiguous codes and originating from accessions which were most likely not apomictic. *Capsella* was used as outgroup, other genera of the Boechereae, *Boechera missouriensisllaevigata*, *Boechera canadensis* and *Boechera repanda* are in the left lower corner, ITS types from *Boechera* species centred in western North America are sorted alphabetically by the name of the ITS type. Taxa for which it was not clear if they were apomictic or sexual are indicated in red

The phylogenetic reconstruction including all ITS-sequence detected in this study is included in the supplementary material on the DVD accompanying the thesis.

Phylogenetic Reconstruction based on AT2G25920.— In the phylogenetic reconstruction based on the AT2G25920 intron *Boechera laevigata*, *Boechera missouriensis*, *Boechera canadensis* and *Boechera repanda* were found on a polytomy together with a large group containing *Sandbergia* and *Boechera* species centred in western North America. *Sandbergia* was sister to the large *Boechera* group. A small group containing accessions assigned to *Boechera stricta*, *Boechera lemmonii, Boechera puberula* and *Boechera lasiocarpa* was placed in sister position to the remaining *Boechera* accessions (figure 2).
The resolution among taxa within the western North American *Boechera* group was poor. In total 19 lineages containing more than one accession were found. 76 additional AT2G25920 sequence types were unresolved on the polytomy. As in the phylogenetic reconstruction based on the ITS some species-specific lineages were recovered while other lineages represented several taxa. Taxa representing own lineages were *Boechera microphylla, Boechera lemmonii, Boechera cobrensis, Boechera glaucovalvula, Boechera schistacea, Boechera stricta, Boechera suffrutescens* (shared with one individual of *Boechera constacei* and *Boechera koehleri*), *Boechera rectissima* (together with one *Boechera arcuata* accession) and *Boechera crandallii.*

Phylogenetic reconstruction based on AT3G18900.— The phylogenetic reconstruction based on the AT3G18900 intron placed *Boechera repanda* as sister to the remaining *Boechera* accessions and *Sandbergia*. The *Boechera* accessions were split into two groups, *Sandbergia* being sister to one of the groups (figure 3). Relationships in the group being sister to *Sandbergia* remained unresolved. On the polytomy a specific lineage for *Boechera lemmonii*, *Boechera lyallii* and *Boechera stricta* was found. Another lineage contained accessions representing *Boechera koehleri, Boechera breweri, Boechera sparsiflora* but also *Boechera microphylla* and *Boechera microphylla* var. *macounii*. AT3G18900 intron types specific to *Boechera crandallii* and *Boechera cobrensis* were exclusively found within the *Boechera* lineage being sister to *Sandbergia*. However they were found to be unresolved on the polytomy among accessions representing *B. perennans, B. pallidifolia, B. williamsii, B. rectissima, B. selbyi, B. lignifera, B. pallidifolia* and *B. schistacea*.
Boechera missouriensis and *Boechera laevigata* were sister to the second clade recovered by the AT3G18900 based phylogenetic reconstruction. *Boechera canadensis* was sister to the remaining *Boechera* accessions in the second clade. *Boechera rigidissima, Boechera constancei* and *Boechera suffrutescens* formed a sister clade to a polytomy from which five lineages arose. The first lineage was a single *Boechera macounii* accession, the second lineage contained *Boechera pendu-*

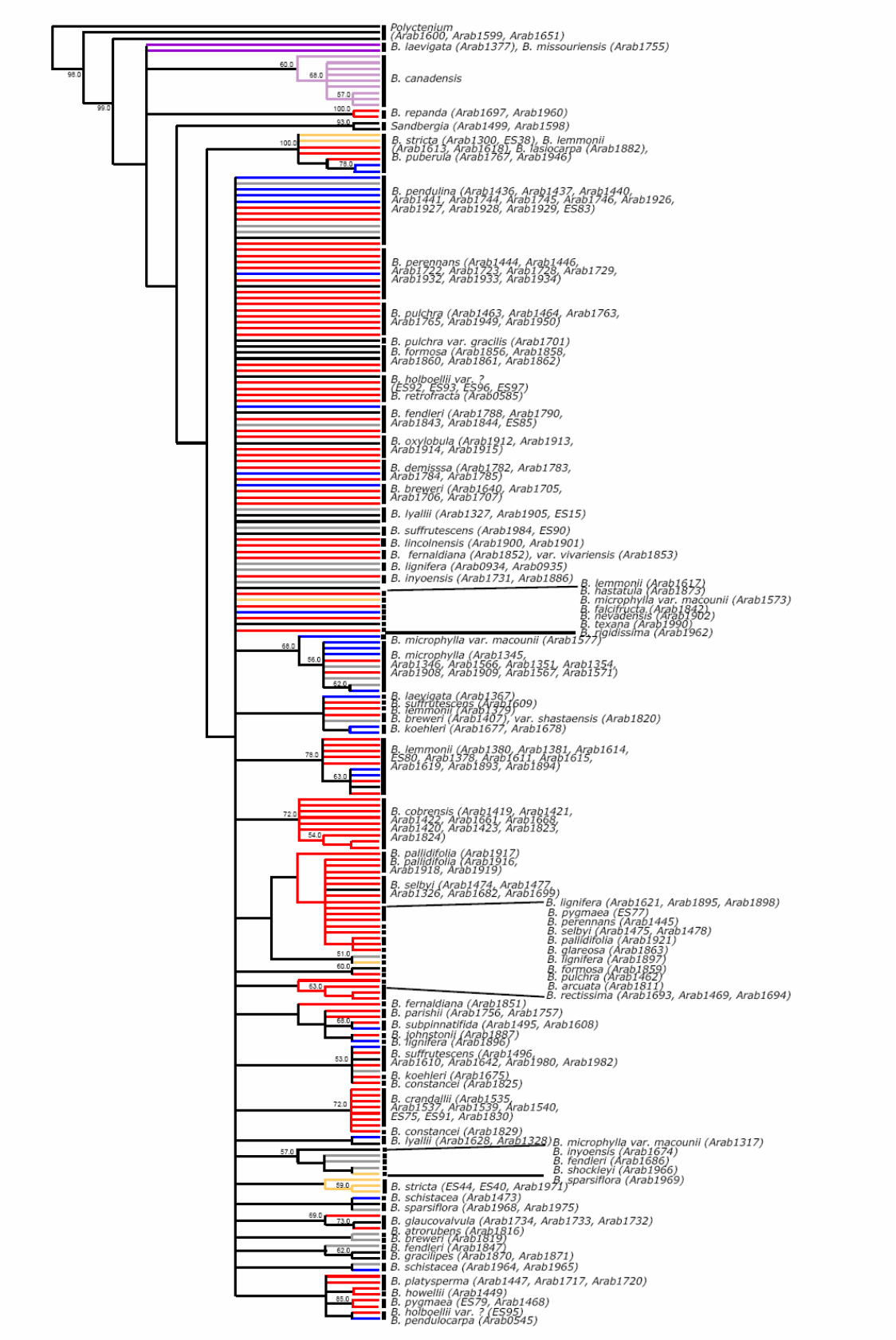

Figure 2 Parsimony analysis of the AT2G25920 intron. *Polyctenium* was used as outgroup. The colours of the branches indicate the lineage of the cpDNA haplotype that was found in this accession (Kiefer et al. submitted); lineage I blue, lineage II yellow, lineage III red, lineage IV pink, lineage V purple, central haplotype and directly derived grey

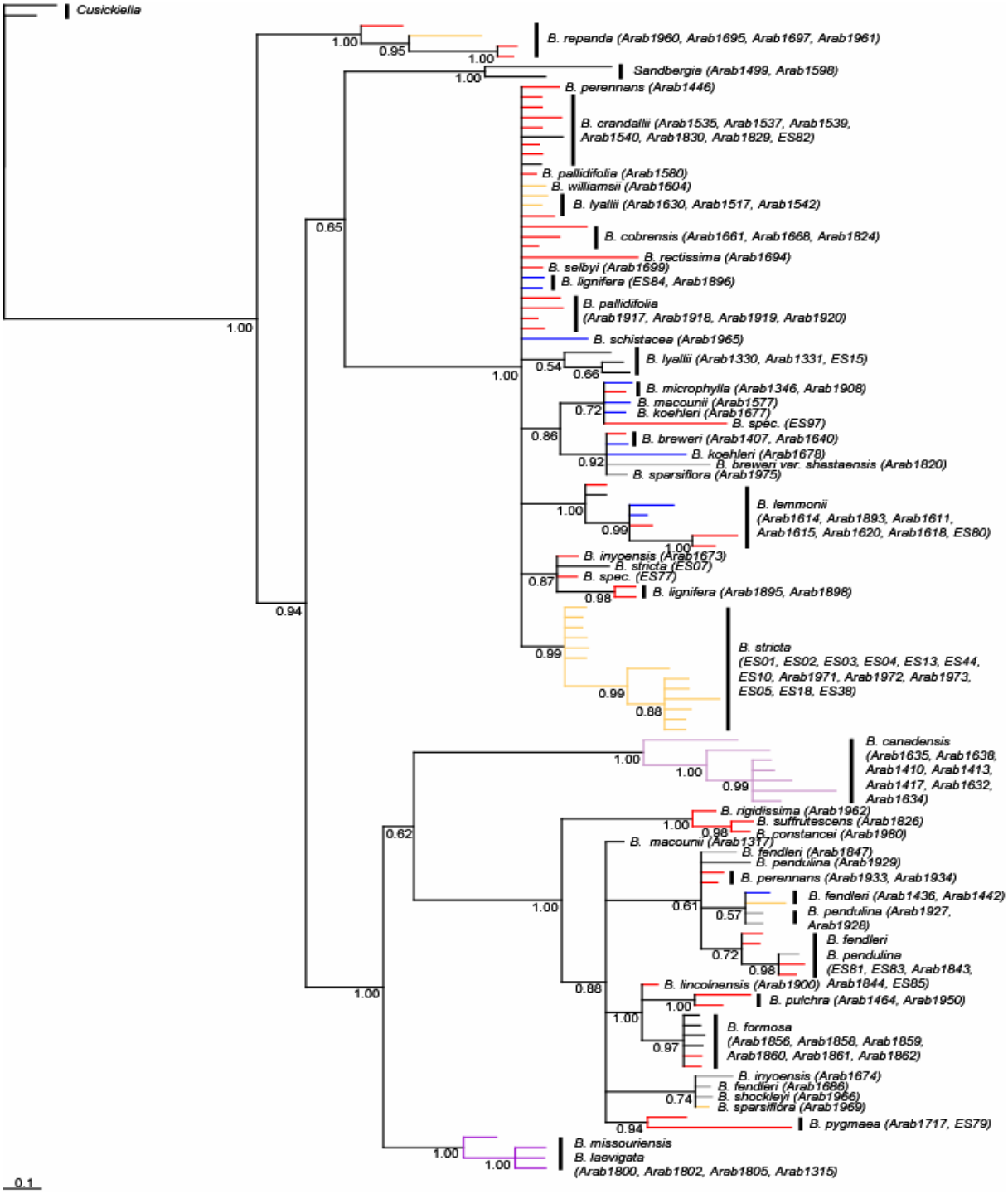

Figure 3 Bayesian analysis of the AT3G18900 intron run through the program TOPALi (Milne et al. 2004). *Cusickiella* was used as outgroup. The colours of the branches indicate the lineage of the cpDNA haplotype that was found in this accession (Kiefer et al. submitted); lineage I blue, lineage II yellow, lineage III red, lineage IV pink, lineage V purple, central haplotype and directly derived grey. Numbers indicate posterior probabilities

lina (six accessions), *Boechera fendleri* (five) and *Boechera perennans* (two accessions).

The third lineage contained six accessions representing *Boechera formosa*, two *Boechera pulchra* and one *Boechera lincolnensis* accession. The fourth lineage

contained one accession of *Boechera inyoensis*, *Boechera fendleri*, *Boechera shockleyi*, and *Boechera sparsiflora*. The fifth lineage contained two *Boechera pygmaea* accessions.

Combined analysis of ITS, AT3G18900 intron and AT2G25920 intron.— The combined analysis resulted in a completely unresolved polytomy (data not shown).

Network analysis of the ITS dataset.— The network analysis was run using the same alignment as for the phylogenetic reconstruction based on ITS data with the non-*Boechera* taxa and eastern North American *Boechera* taxa being omitted. This analysis resulted into a network with a mixture of clearly separated lineages (ab/bv, er/ly/lz, ks/hl/gy/kr, pq/la/lb/lc/po/pu, rb/pp arising from ITS type ad; pg/ph/pi/lk and na/nb/nc arising from ie which is connected to ad via ee; ar/gx/pf/sa/kn/mm/ha/mv connected to h; nh/ky/su/rl/ld connected to h) and reticulate structures (e.g. op/ml/mn/mo/ie/hm/ee connected to ad) (figure 4). Especially the ITS types ac, v, g and f all themselves being origin for separate lineages and connected to h could not be resolved. Most of the reticulate pattern was caused by repeated mutations in alignment position 216 and 444. Hence, the alignment was subdivided into sets of ITS-types according to the phylogenetic reconstruction based on ITS sequences. ITS type ac and v gave both rise to four separate lineages and were connected to h by two mutation steps (figure 5a and 5b). ITS type g was connected to ITS type h by one mutation step and was origin for four lineages. ITS type bt which was connected to g by one mutation step was also connected to a series of missing ITS types which connected a group of four ITS types to the network (z/mp/le and og/sx). Those two groups of ITS types were also connected to ITS type h by several missing ITS types (figure 5c). Finally ITS-type f was connected to ITS type h by two mutation steps and gave rise to 11 lineages (figure 5d).

Geographic distribution of subgroups.— Distribution ranges of most ITS-types overlapped. However, for the most frequent ITS-types a centre of distribution could be determined (Table 1). ac, au, eu, f, g, gx and op had had their major occurrence in the Great Basin and on the Snake/Columbia Plateau. ab, bt, l and t had their distribution centre in the Rocky Mountains. c and r occurred to the same extent in the Great Basin or Snake/Columbia Plateau and the Rocky Mountains. az was found in the Rocky Mountains, the Cascades and in general north of the last glaciation. e was exclusively found north of the last glaciation. ev was typically found in the deserts as well as on the Colorado Plateau. z had its centre of distribution in the Sierra Nevada. h and ad were both central to the network analysis and both had several lineages arising from them. h had a continuous distribution range but it occurred mostly in the Great Basin, the Rocky Mountains and also north of the last glaciation. ad had a disjunct distribution range and was mainly present in the western Great Basin (mainly

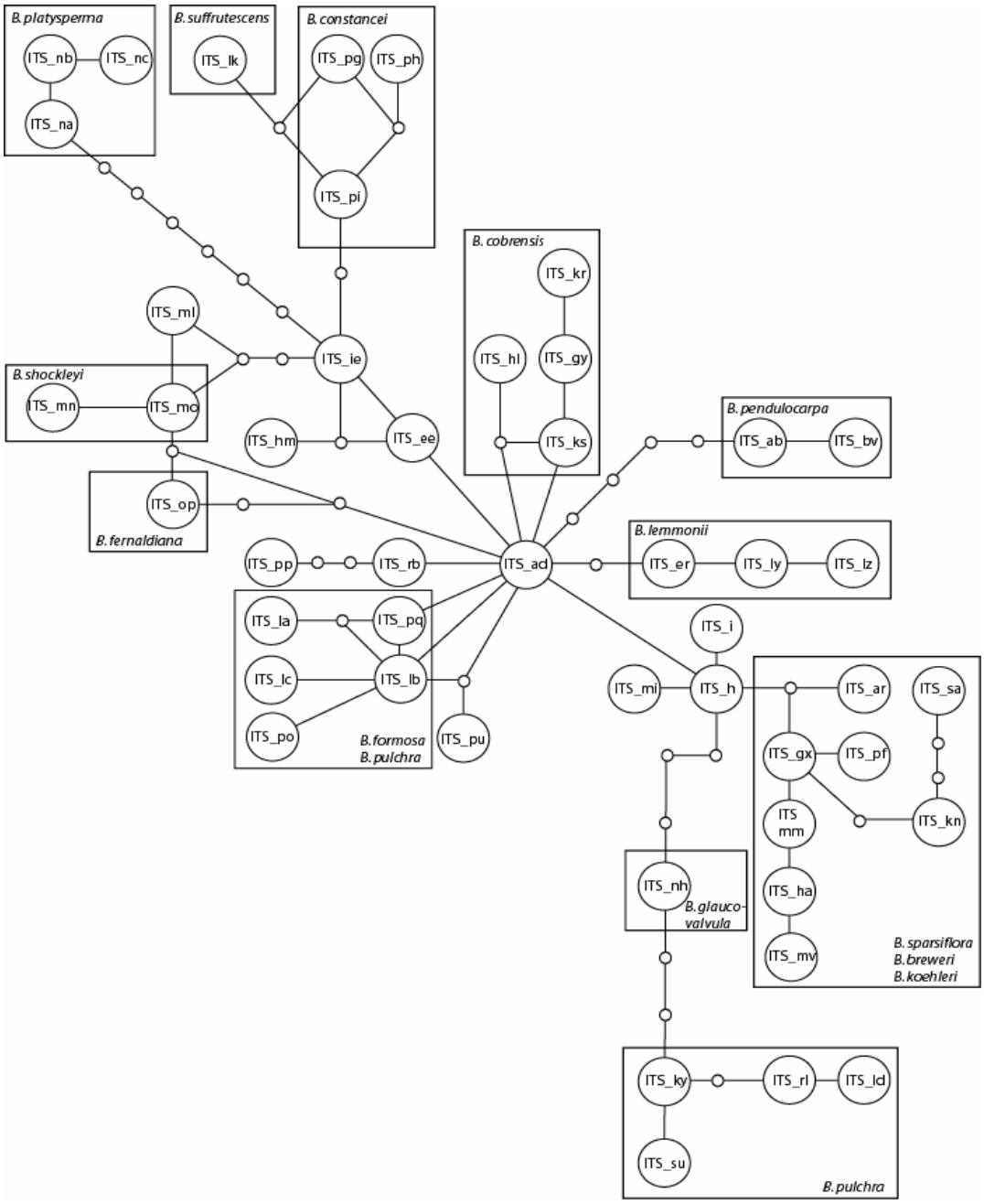

Figure 4 Parsimony network analysis of the analysis of the alignment reduced to sequences without ambiguous codes except N and sequences originating only from accessions which were most likely not apomictic according to their taxonomic determination; only the resolved section of the network is shown. Species names in the squares indicate the taxa which predominantly shared the ITS types included in the squares

Boechera cobrensis) and on the Colorado Plateau (mainly *Boechera pallidifolia* and *Boechera selbyi* = *Boechera gracilenta*).

Three of the subgroups from the network analysis were analysed in greater geographic detail. ITS type ac was central to the first group (figure 6a). It had a disjunct distribution range and was found in Klamath/Siskiyou (*Boechera subpinnatifida* and *Boechera koehleri*), the western Great Basin (*Boechera*

retrofracta and *Boechera pinetorum*), the northern Wasatch/Uinta Range (*Boechera lignifera* and *Boechera retrofracta*) and the southern central Rockies (*Boechera retrofracta*). Derived from it were ITS types eu and ru in Klamath Siskiyou (*Boechera subpinnatifida*) and in the western Great Basin and on the western Snake/Columbia Plateau (*Boechera puberula*, *Boechera pulchra*, *Boechera retrofracta*). ITS type au was also derived from ITS type ac and occurred in the western half of the Great Basin (*Boechera puberula* and others). ITS type oh also derived from ITS type ac occurred in the northern Wasatch/Uinta range (*Boechera lasiocarpa*).

ITS-type g was derived from ITS type h by one mutation step. It was mainly found in *Boechera pendulina* individuals in the central Great Basin and the Wasatch Uinta Range (figure 6b). However, moving further southeast into the southern Wasatch/Uinta Range, on the Colorado Plateau and the Arizona Mountains, ITS type g was detected in *Boechera fendleri* individuals. The lineage containing ITS types is and ke was derived from ITS type g. Those ITS types were restricted to *Boechera* fendleri in the Mojave Desert. ITS types kd, rd and gz were also derived from ITS type g. They had a circular connection to ITS type g, so it is unclear if they all belong to the same lineage or if they are individually derived from ITS type g. They were found in *Boechera fendleri* (one individual), *Boechera pendulina* (one individual) and also in three other species (three individuals) on the northern Colorado Plateau. ITS type ho was derived from ITS type g and found in the Chihuahua Desert in one individual of *Boechera* fendleri. Another larger lineage arising from ITS type g was dominated by *Boechera pendulocarpa* individuals (bt, bw, bx, bu, eq). Those ITS types had a scattered distribution in the western and north eastern Great Basin, Sierra Nevada, Klamath/Siskiyou, the Colorado Rockies, Okanogan Forest, Eastern Cascades, Montana Valley Foothill Grasslands and Wasatch/Uinta Range with a focus in the southern central Rockies.

ITS type gx was derived from ITS type h by two mutation steps and was origin to a lineage of mainly *Boechera sparsiflora*, *Boechera breweri* and *Boechera koehleri* individuals (figure 6c). ITS type gx was dominated by *Boechera sparsiflora* individuals. It occurred along the western border of the Great Basin, reaching into the Northern Sierra Nevada and the northwestern Snake/Columbia Plateau. In one accession of *Boechera breweri* it occurred also in Klamath/Siskiyou. ITS types mm, ha and mv as well as ITS types kn and na, and pf made up three lineages derived from ITS type gx. They all represented a mixture of *Boechera* breweri and *Boechera koehleri* individuals (along with one *Boechera retrofracta* accession) in Klamath/Siskiyou and the Northern Sierra Nevada.

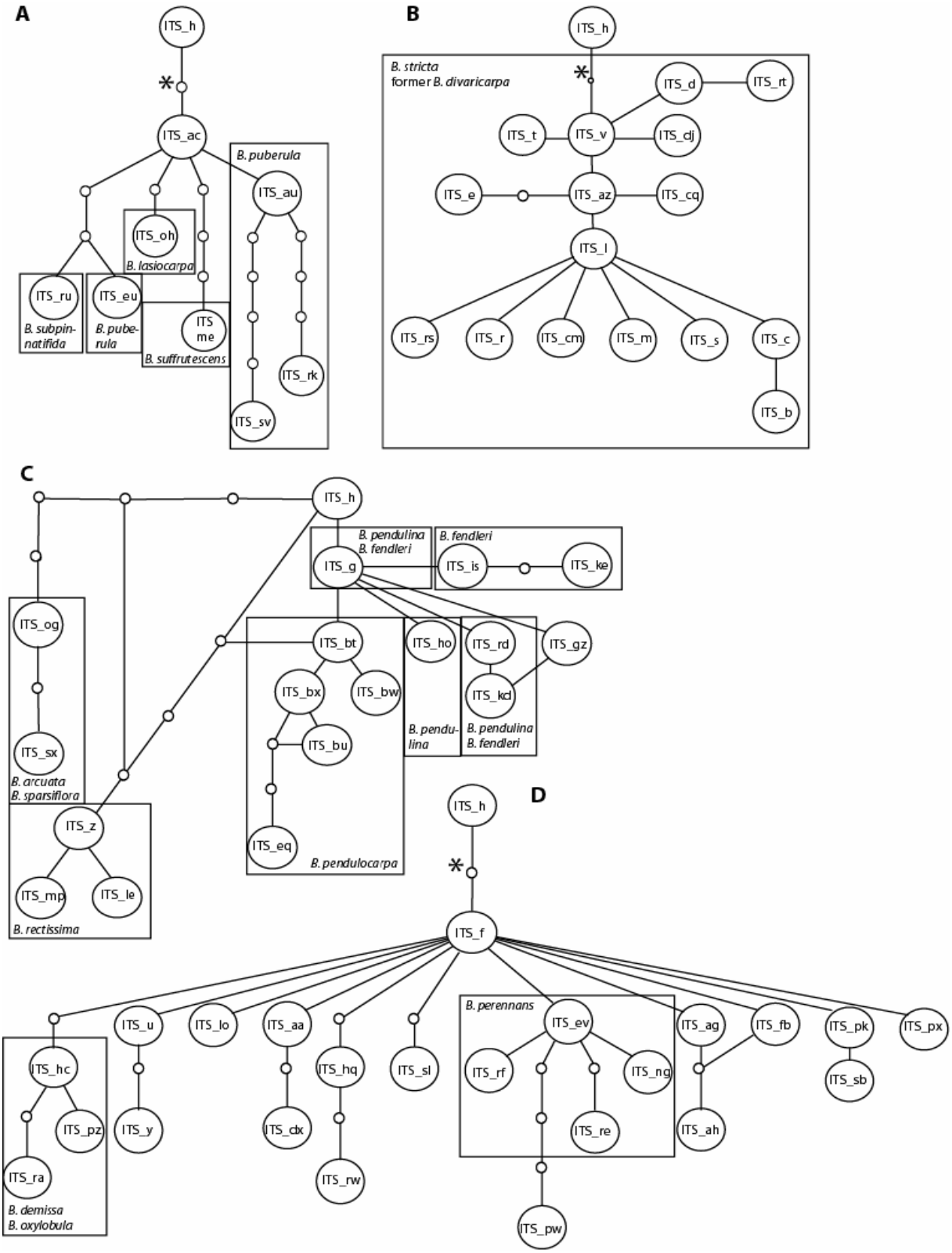

Figure 5 Parsimony network analysis based on ITS sequences; subgroups were defined by the groups recovered in the parsimony analysis. ITS type h was added to all analysis for rooting; A ITS type ac and derived lineages, B *Boechera stricta* specific lineage, C ITS type g and derived lineages with additional ITS types that have a circular connection to a lineage arising from ITS type g, D ITS type f and derived lineages; species names in the squares indicate the taxa which predominantly shared the ITS types included in the squares

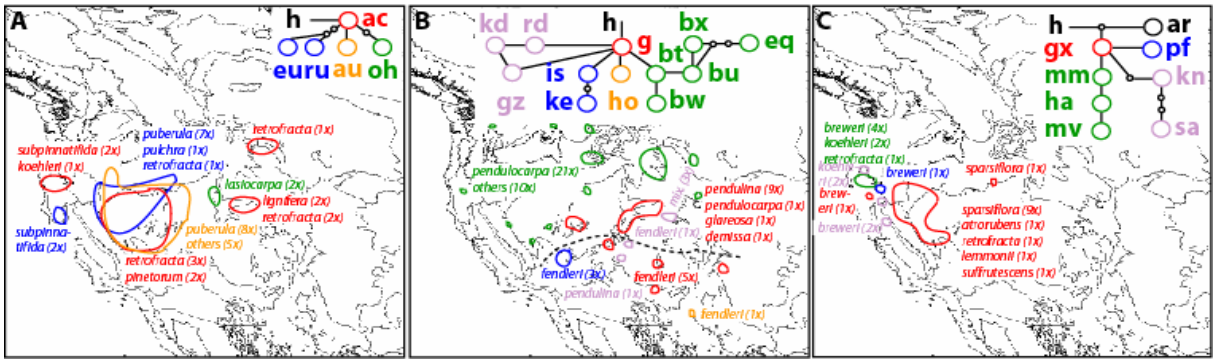

Figure 6 Geographical distribution ranges of selected ITS lineages; in each figure a simplified version of the network section is given and colour coded according to the distribution ranges indicated in the map. Next to each distribution range a list of the species sharing the ITS type in the encircled region is given. The number in brackets indicates the number of individuals; A distribution range of ITS type ac and derived lineages, B distribution range of ITS type g and derived lineages, C distribution range of ITS type gx and derived lineages

Divergence of sequences within subgroups.— Divergence of sequences within subgroups defined by the phylogenetic reconstruction based on the ITS dataset was calculated as average p-distances (Table 2). The core *Boechera* group (all *Boechera* species included in this study centred in western North America had an average divergence of 0.011. All subgroups had an average divergence between 0,002 and 0.005.

Assuming a mutation rate for the ITS of approximately 0.5% to 2.5% nucleotide divergence per 1 million years (for comparison: highest substitution rates with 5.3 3 10–9 substitutions/site/year, Wendel, Schnabel, and Seelanan 1995; lowest rates with 0.35 3 1029 substitutions/site/year, Suh et al. 1993) a p-distance of 0.01 correspond to 1 million years (as a mean). It is obvious that this calculation is highly biased, and it is used herein simply as a very rough indicator.

Comparison of cpDNA-types, ITS types and single copy allele identity.— Phylogenetic reconstruction based on nrDNA ITS, AT2G25920 and AT3G18900 revealed lineages specific to some species or a group of species as well as putative hybrids which showed up among specific lineages. ITS-types and intron sequence types as well as their position in network and phylogenetic reconstruction were compared for additional information on the nature of a specific lineage or hybrid. For information of the maternal side cpDNA haplotypes as defined earlier (Kiefer et al. submitted) were also included in the comparison (supplementary table 7).

The comparison revealed that some lineages were uniform and recovered at least in the phylogenetic reconstruction based on two of the marker sequences while other species had a complex mixture of sequence types from different lineages.

Table 1 Distribution ranges of major ITS types; numbers indicate the number of individuals carrying this ITS type in the named region. Bold numbers indicate the centres of distribution defined by number of individuals.

ITS type	Klamath Siski-you	Great Basin Snake	Sierra Neva-da	Des-erts	Rocky Mtns.	Cas-cades	N of glaciation	Wasatch Uinta	Colo-rado Pla-teau	other
ab	-	-	-	-	**4**	-	1	3	-	-
ac	3	**6**	-	-	3	-	-	-	-	-
ad	1	**12**	1	-	1	3	-	4	**12**	-
ar	-	2	2	-	-	1	-	-	-	-
au	-	**13**	-	-	-	-	-	-	-	-
az	-	1	1	-	**6**	**7**	**5**	-	-	-
bt	1	3	1	-	**12**	-	1	-	-	1
bx	-	-	-	-	2	1	-	1	-	2
c	-	**6**	-	-	**5**	-	-	3	2	-
e	-	-	-	-	-	-	**38**	-	-	-
eu	-	**8**	-	-	-	1	-	-	-	-
ev	-	4	1	**9**	-	-	-	-	**8**	1
f	-	**8**	-	3	2	-	-	1	4	3
g	-	**6**	-	2	-	-	-	4	3	2
gx	1	**9**	2	-	-	2	-	-	-	1
h	-	**30**	8	-	**36**	7	**25**	5	-	10
i	**6**	1	**6**	-	-	-	-	-	-	-
l	-	-	-	-	**25**	-	2	1	2	1
op	-	**8**	-	-	-	-	-	-	-	-
r	-	**7**	-	-	**5**	-	-	2	-	-
t	2	-	1	-	**8**	3	-	-	1	1
v	-	3	2	-	4	-	-	2	-	-
z	1	-	**6**	-	1	-	-	-	-	1

Table 2 Divergence time estimates of subgroups calculated from average sequence divergence calculated through the program TOPALi (Milne et al. 2004). The first column gives the most frequent ITS type of a group, the second column gives the taxa dominating this lineage, the third column gives the average pairwise distances and the fourth column gives the age estimate assuming an average divergence of 1% per 1 million years (Koch et al. 2003).

phylogroup	predominant taxa	average pairwise distance	age estimate
all core *Boechera*	-	0.0111	1 my
ac group	*B. puberula*	0.0045	450,000 y
az group	*B. stricta*	0,0042	420,000 y
bt group	*B. pendulocarpa*	0,0035	350,000 y
ev group (network)	*B. perennans*	0,0029	290,000 y
f group (network)	-	0,0051	510,000 y
gx group	*B. sparsiflora* *B. breweri* *B. koehleri*	0,0035	350,000 y
kv group	*B. platysperma* *B. rigidissima*	0,0083	830,000 y
ky group	*B. pulchra*	0,0044	440,000 y
le group	*B. arcuata* *B. sparsiflora* *B. rectissima*	0,0050	500,000 y
ml group	*B. shockleyi* *B. inyoensis*	0,0031	310,000 y

Boechera cobrensis shared ITS type ad and ITS-types derived from it. In the phylogenetic reconstruction based on the AT2G25920 intron the accessions also clustered in one lineage. AT3G18900 was sequenced for three of the accessions and put them together in one of the recovered lineages but in an unresolved relationship. The cpDNA haplotype was exclusively haplotype ci, the most frequent cpDNA type in the cpDNA analysis (Kiefer et al. submitted).
If *Boechera fendleri* shared ITS type g it was unresolved in the analysis based on AT2G25920 intron but occurred in one lineage in the phylogenetic reconstruction based on the AT3G18900 intron. The accession which shared ITS type f was also unresolved in the analysis based on the AT2G25920 intron. The accession which carried ITS type sn showed up in a group of mixed species which was recovered in the phylogenetic reconstruction based on the AT2G25920 intron as well as on the AT3G18900 intron. This individual also carried a haplotype from the evolutionary chloroplast lineage II (Kiefer et al. submitted). *Boechera formosa* was also an example for a simple lineage. The ITS types were either ad or derived from ad, in AT2G25920 the relationship of the sequences representing these accessions was unresolved and in the phylogenetic reconstruction based on AT3G18900 they constituted a separate lineage within clade 2.
Boechera lemmonii was interesting in so far that several individuals were found to be in a separate lineage in the phylogenetic reconstructions based on either one of the employed markers. However, two accessions carried cpDNA haplotype from lineage I unlike the other accessions which all carried haplotypes from chloroplast lineage III (Kiefer et al. submitted). One individual which carried a different ITS type showed up in a lineage of mixed species in the phylogenetic reconstruction based on AT2G25920. Three other individuals were found in one ITS-lineage but showed up in different lineages in the phylogenetic reconstruction based on the AT2G25920 intron. However, if the chloroplast DNA haplotype was known it was found to be in chloroplast lineage III.
Boechera lignifera had a wild mixture of ITS types from different lineages, among them sequences with ambiguous codes. Also in the phylogenetic reconstruction based on the AT2G25920 intron and the AT3G18900 intron the accessions showed up in different positions.
Boechera microphylla (in the sense of Rollins 1993) shared ITS type h mainly with *Boechera retrofracta*. Commonly *Boechera microphylla* accessions shared chloroplast haplotypes from lineage I. In the phylogenetic reconstruction based on the AT2G25920 intron *Boechera microphylla* accessions were found in a separate lineage except for two individuals which showed up in a lineage of mixed species or unresolved. The individual carrying the AT2G25920 allele whose relationship was unresolved carried a chloroplast haplotype from lineage III. The individual whose AT2G25920 allele was found in the mixed lineage showed up in an unresolved position in clade 2 of the phylogenetic reconstruction based on the AT3G18900 intron

while other accessions which carried ITS type h and were in a separate lineage based on the AT2G25920 intron were found in clade on in a lineage with *Boechera sparsiflora*, *Boechera breweri* and *Boechera koehleri* in clade 1.
Boechera pallidifolia consistently showed up together with *Boechera selbyi* (according to Al-Shehbaz unpublished = *Boechera gracilenta*).
Boechera suffrutescens showed up consistently with *Boechera constacei* also when the chloroplast DNA haplotype was from different lineages.
Boechera perennans, *Boechera pendulina* and *Boechera fendleri* had a complex relationship with each other. Typical for *Boechera pendulina* and *Boechera fendleri* was ITS type g. However it also carried ITS type ev in a small region on the western Colorado Plateau. *Boechera perennans* mainly carried ITS type ev. *Boechera pendulina* mainly carried chloroplast lineage I haplotypes while *Boechera perennans* typically carried lineage III haplotypes. However, *Boechera perennans*, *Boechera pendulina* and *Boechera fendleri* showed up together in one lineage in clade 2 in the analysis of AT3G18900 data.

DISCUSSION

We based our phylogenetic reconstructions on three different marker systems all with different properties or locations: (a) nrDNA ITS which is present in the genome in thousands of copies and is subject to concerted evolution, a fact which may cause trouble in defining true lineages, (b) an intron of AT3G18900, a single copy gene on *Boechera stricta* chromosome 3 not undergoing concerted evolution, and (c) an intron of AT2G25920, a single copy gene on *Boechera stricta* chromosome 5 not undergoing concerted evolution. Neither the analysis of ITS1 and ITS2 nor the analysis of AT2G25920 resulted into a well resolved tree. The analysis of AT3G18900 resulted into two clades which were not or only little resolved.
An earlier study of nrDNA in *Boechera* including *Boechera stricta*, *Boechera holboellii* sensu Rollins (Rollins 1993) and *Boechera divaricarpa* sensu Rollins (Rollins 1993) had already shown that there was no resolution among ITS types and only few ITS types were joined in separate groups (Dobeš et al. 2004b). Unresolved trees are a common phenomenon, usually regarded as experimental failure (Rokas and Carroll 2006). However, the lack of resolution indicates that the time span between branching events was extremely small which is a useful information. If a polytomy is caused by truly simultaneous cladogenesis it is referred to as a hard polytomy while a polytomy caused by too few characters and superimposed substitutions it is called a soft polytomy (Maddison, 1989). Since the polytomy was recovered for ITS and AT2G25920 in parallel and also in the phylogenetic reconstruction based on AT3G18900 internal branches are shorter than terminal branches the polytomy detected in our dataset can be regarded as a hard polytomy showing the rapid

cladogenesis in *Boechera*. Many of the lineages we recovered were species specific. Therefore, the shape of the trees cannot only be interpreted as rapid cladogenesis but indeed as rapid speciation.

Every phylogenetic reconstruction based on sequence data does not necessarily represent the phylogenetic history of the species. It only shows relationships among the different alleles of the piece of DNA under investigation. However, if several DNA sequences support the same tree topology it is likely sequence evolution and species evolution are congruent in this case. With so many unresolved lineages it is hard to find congruent topologies. Nevertheless we found lineages which were recovered two or all of the phylogenetic reconstructions. However, often in one dataset a lineage was resolved while it was unresolved in the other. If a “good lineage” could be identified and only few individuals showed up among other lineages or had ambiguous sites in their ITS sequence these were interpreted as hybrids. That way we could identify hybrids for example in *Boechera fendleri*, *Boechera lemmonii*, *Boechera microphylla*, *Boechera perennans*, and *Boechera pendulina*.

The nrDNA ITS dataset was also subjected to a network analysis. The network was found to have two central ITS types, ad and h. From both centres several species specific or mixed lineages originated. The lineages originating from ITS type ad were mostly without conflicting data and circular connections were rare. On the other hand connections to lineages originating from ITS type h were often in conflict with each other. These conflicts were mainly due to two alignment positions, namely 216 and 444. *Boechera stricta* typically had the combination A in 216 and T in 444. The majority of accessions had the combination G in 216 and T in 444. This means that position 216 can be used to tell apart the *Boechera stricta* lineage from the other lineages. Hence it follows that a group with A in position 216 is derived from the *Boechera stricta* lineage, such as the lineage including ITS type f. On the other hand a group with G in position 216 is derived from the majority, such as the lineage including ITS type g. The lineages including ITS types f and g both have G in position 444. There are three possibilities which may explain this. First of all a mutation might have occurred in parallel in position 444 changing T into G in parallel. The second possibility is that in the lineage including most accessions the mutation in position 444 happened first giving rise to ITS type g. Later hybridisation with a *Boechera stricta* individual and concerted evolution may have produced ITS type f. Both possibilities are equally parsimonious because they involve two steps. However, *Boechera stricta* is also characterized by a mutation in position 161 which is neither present in ITS type f or g or any of the ITS types derived from them. Hence, it is most likely that the mutation in position 444 occurred in parallel in two lineages.

ITS lineages in which the central ITS type was shared by several species were also analysed geographically. ITS type ac was such a case. It showed a fragmented distribution in Klamath/Siskiyou, the western Great Basin, the northern Wasatch/Uinta Range and with one accession in the southern central Rockies. Interestingly ITS type ac was mainly found in *Boechera retrofracta* individuals. Two lineages dominated by *Boechera puberula* split from ITS type ac in the western Great Basin and on the Snake/Columbia Plateau co-occurring with ITS type ac. In Klamath/Siskiyou ITS type ac was shared by *Boechera subpinnatifida* (two accessions) and *Boechera koehleri* (one accession). *Boechera subpinnatifida* also carried ITS type ru derived from one of the *Boechera puberula* types in Klamath/Siskiyou. It seems possible that either the distribution range of ac and eu used to reach Klamath/Siskiyou in the past and after range diffraction *Boechera* subpinnatifida evolved in this ecoregion. However, it could also be that the distribution range of ac was never larger and that ITS types ac and eu were carried into Klamath/Siskiyou by the widespread *Boechera retrofracta*. Since we do not have information about the speciation process itself we cannot rule out one of the possibilities.

Another interesting example is ITS type g and its derived lineages. ITS type g is typical for *Boechera pendulina*. It occurs in the central Great Basin, the Wasatch/Uinta Range, on the Colorado Plateau and in the Arizona Mountains. Remarkably there seems to be a boundary to the south beyond which ITS type g is found in *Boechera fendleri*. Additionally from three other separate lineages originating from ITS type g two are specific to *Boechera fendleri* and only in the southern range. The third lineage includes a mix of *Boechera fendleri*, *Boechera pendulina* and other species. North of the distribution range of ITS type g lies the distribution range of a lineage dominated by *Boechera pendulocarpa. So* unlike in the lineage including ITS type ac it is not range fragmentation which may have played a role in speciation but rather range expansion to the north and south.

Last but not least ITS type gx, also derived from ITS type h, is a second example for a lineage where species differentiated in the Klamath/Siskiyou ecoregion. ITS type gx is dominated by *Boechera sparsiflora* and is distributed along the most western part of the Great Basin reaching into the Sierra Nevada and into Klamath/Siskiyou with one *Boechera breweri* accession. The three lineages derived from ITS type gx occur only in the Klamath/Siskiyou and the Sierra Nevada and with one exception are shared between *Boechera koehleri* and *Boechera breweri*.

Since ITS type gx is derived from ITS type h and ITS type h is absent from Klamath/Siskiyou it seems likely that first *Boechera sparsiflora* differentiated in the western Great Basin from the ancestor carrying ITS type h followed by migration or more likely dispersal into Klamath/Siskiyou where *Boechera koehleri* and *Boechera breweri* evolved. Alternatively the distribution range of *Boechera breweri* and *Boechera koehleri* was larger in the past reaching into the western Great Basin. Then

glaciation of the Sierra Nevada and Southern Cascades separated the unglaciated Klamath Siskiyou from the unglaciated Great Basin and lead to the extinction of the two species in this range.

Average sequence divergence of ITS types allowed for a rough age estimate. Taking the average value of 1% sequence divergence per one million years (Koch et al. 2003) the age of ITS lineages representing species centred in western North America is one 1my. The age of the identified lineages is between 500,000 and 200,000 years. This time falls entirely into the second half of the quaternary. During this time at least three glaciation cycles took place (Bintanja and van der Wal, 2008). The quaternary ice ages forced plant and animal taxa to migrate into and out of glacial refugia (Hewitt, 2000). Changing climate promoted the formation of new habitats which supported speciation (Hewitt, 2004, Willis and Niklas, 2004). Regarding the time frame in which the ITS lineages developed makes it likely that *Boechera*'s diversity as we find it today is also a result of those climatic oscillations.

SUMMARY

The phylogeny of *Boechera* could neither be resolved by nrDNA ITS nor by sequencing two introns from single copy genes. Species specific lineages could be detected but it cannot be determined which lineages are sister to each other. Interpreting the structure of the trees with either unresolved lineages in the case of ITS and the AT2G25920 intron or short internal branches as in the case of AT3G18900 indicates rapid cladogenesis and in this case rapid speciation.
Even if *Boechera* underwent radiation the split of lineages was successive. To unravel those quick processes is not possible with sequencing marker sequences. A continental wide population genetics approach might be the most promising approach.

LITERATURE CITED

AL-SHEHBAZ, I. A. (submitted) *Boechera* A. Löve and D. Löve. In: Flora North America.

AL-SHEHBAZ, I. A. 2003. Transfer of most North American species of *Arabis* to *Boechera* (Brassicaceae). Novon 13: 381-391.

AL-SHEHBAZ I. A., BEILSTEIN, M. A., AND E. A. KELLOGG. 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Systematics and Evolution 259: 89–120.

AL-SHEHBAZ, I. A. 2007. The North American genus *Sandbergia* (Boechereae, Brassicaceae). Harvard Papers of Botany 12: 425–427.

BAILEY, C. D., KOCH, M. A., MAYER, M., MUMMENHOFF, K., O'KANE, S. L., WARWICK, S. I., WINDHAM, M. D., AND I. A. AL-SHEHBAZ. 2006. A Global nrDNA ITS Phylogeny of the Brassicaceae. Molecular Biology and Evolution 23: 2142–2160.

BINTANJA, R., AND R. S. W. VAN DER WAL 2008. North American Ice Sheet Dynamics and the onset of 100,000 year glacial cycles. Nature 454: 869–872.

BÖCHER, T. W. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Kong Danske Vidensk Selsk, 6: 1–59.

CLEMENT, M., POSADA, D., AND K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9: 1657–1659.

DOBEŠ, C., MITCHELL-OLDS, T. AND M. A. KOCH. 2004a. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* ×*divaricarpa*, and *A. holboellii* (Brassicaceae). Molecular Ecology 13: 349–370.

DOBEŠ, C., MITCHELL-OLDS, T. AND M. A. KOCH. 2004b. Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera x divaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. American Journal of Botany 91: 2087–2101.

DOBEŠ, C., KOCH, M.A., AND T. F. SHARBEL. 2006. Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Annals of the Missouri Botanical Gardens 93: 517–533.

DOBEŠ C., KOCH M., AND C. KIEFER. 2007. Apomixis and radiation at low ploidy levels exemplified in the evolutionary model genus *Boechera* (Brassicaceae). Regnum Vegetabile 147: 391–412.

DOBEŠ, C., SHARBEL, T. F, AND M. A. KOCH. 2007. Towards understanding the dynamics of hybridization and apomixis in the evolution of genus *Boechera* (Brassicaceae). Systematics and Biodiversity 5: 321–332.

DOYLE, J. J. AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

ELDER J. F., AND B. J. TURNER. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Quarterly Reviews of Biology 70: 297–320.

FELINER G. N., LARENA B. G., AND J. F. AGUILAR. 2004. Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). Annals of Botany 93: 189–200.

HAZEN, S. P., BOREVITZ, J. O., HARMON, F. G., PRUNEDA-PAZ, J. L., SCHULTZ, T. F., YANOVSKY, M. J., LILJEGREN, S. J., ECKER, J. R., AND STEVE A. KAY 2005. Rapid Array Mapping of Circadian Clock and Developmental Mutations in Arabidopsis. Plant Physiology 138: 990–997.

HEWITT, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405: 907–913.

HEWITT, G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transaction of the Royal Society of London B. Biological Sciences 359:183–95.

HORNEMANN, J. W. 1827. Flora Danica. 11 (32): 1879.

KIEFER, C., DOBEŠ, C., SHARBEL, T. F., M. KOCH. (submitted) Phylogeographic structure of the chloroplast DNA gene pool in North American *Boechera* – a genus and continental wide perspective.

KOCH, M., BISHOP, J. AND T. MITCHELL-OLDS. 1999. Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biology: 1: 529–537.

KOCH, M., HAUBOLD, B., AND T. MITHCHELL-OLDS. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, *Arabis* and related genera. Molecular Biology and Evolution. 17: 1483–1498.

KOCH, M., DOBEŠ, C., AND T. MITCHELL-OLDS. 2003. Multiple hybrid formation in natural populations: Concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molecular Biology and Evolution 20, 338–350.

KOCH, M., KIEFER, C., VOGEL, J., EHRICH, D., BROCHMANN, C., AND K. MUMMENHOFF. 2006. Three times out of Asia Minor - the phylogeography of *Arabis alpina* L. (Brassicaceae). Molecular Ecology 15: 825–839.

LEVY, Y. Y., MESNAGE, S., MYLNE, J. S., GENDALL, A. R., AND C. DEAN. 2002. Multiple Roles of Arabidopsis VRN1 in Vernalization and Flowering Time Control Science: 297(5579): 243–246

LIU, X. L., COVINGTON, M. F., FANKHAUSER, C., CHORY, J., AND D. R. WAGNER. 2001. *ELF3* Encodes a Circadian Clock–Regulated Nuclear Protein That Functions in an Arabidopsis *PHYB* Signal Transduction Pathway; The Plant Cell13: 1293–1304.

LÖVE, A. AND D. LÖVE. 1976. Botaniska Notiser 128: 513.

MADDISON, W. P. 1989. Reconstructing Character Evolution in polytomous cladograms. Cladistics 5: 365–377.

MILNE, I., WRIGHT, F., ROWE, G., MARSHAL, D. F., HUSMEIER, D. AND G. MCGUIRE. 2004. TOPALi: Software for Automatic Identification of Recombinant Sequences within DNA Multiple Alignments, Bioinformatics 20: 1806–1807

ROKAS, A., AND B. C. CARROLL. 2006. Bushes in the tree of life. Public Library of Sciences Biology 4: 1899–1904.

NICHOLAS, K.B. AND H. B. JR. NICHOLAS. 1997. GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author.

ROLLINS, R. C. 1941. A monographic study of *Arabis* in western North America. Rhodora, 43: 289–325 348–411 425–481.

ROLLINS, R.C. 1993. The Cruciferae of Continental North America. Stanford University Press, Stanford.

SCHRANZ, M. E., WINDSOR, A. J., SONG, B.-H., LAWTON-RAUH, A., AND T. MITCHELL-OLDS. 2007. Comparative Genetic Mapping in *Boechera stricta*, a Close Relative of Arabidopsis. Plant Physiology 144: 286–298.

SIENA, L., SARTOR, M., ESPINOZA, F., QUARIN, C., AND J. ORTIZ. 2008. Genetic and embryological evidences of apomixis at the diploid level in paspalum rufum support recurrent auto-polyploidization in the species. Sexual Plant Reproduction 21: 205–215.

SHARBEL, T. F., AND T. MITCHELL-OLDS. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (*Brassicaceae*). Heredity 87: 59–68.

SONG, B. H., CLAUSS, M. J., PEPPER, A. AND T. MITCHELL-OLDS. 2006. Geographic patterns of microsatellite variation in *Boechera stricta*, a close relative of *Arabidopsis*. Molecular Ecology 15: 357–369

SUH, Y., THEIN, L. B., REEVE, H. E., AND E. A. ZIMMER. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. American Journal of Botany 80:1042–1055.

SWOFFORD, D.L. 2002. *PAUP*: Phylogenetic Anaylsis Using Parsimony (*and Other Methods)*, Version 4.0B10. Sinauer Associates, Sunderland, Massachusetts

WARWICK, S. I., FRANCIS, A., AND I. AL-SHEHBAZ. 2006. Brassicaceae species checklist and database on CD-ROM. Plant Systematics and Evolution 259: 249–258.

WENDEL, J. F., A. SCHNABEL, AND T. SEELANAN. 1995. An unusual ribosomal DNA sequence from Gossypium gossypioides reveals ancient, cryptic, intergenomic introgression. Molecular Phylogenetics and Evolution 4:298–313.

WILLIS, K.J., AND K. J. NIKLAS. 2004. The role of quaternary environmental change in plant macroevolution: the exception or the rule? Philosophical Transactions of the Royal Society of London B. 359, 159–172

WINDHAM, M.D. AND I. A. AL-SHEHBAZ. 2006. New and noteworthy species of *Boechera* I: sexual diploids. Harvard Papers of Botany 11: 61–88.

WINDHAM, M.D. AND I. A. AL-SHEHBAZ. 2007a. New and noteworthy species of *Boechera* (Brassicaceae) II: apomictic hybrids. Harvard Papers of Botany 11: 257–274.

WINDHAM, M.D. AND I. A. AL-SHEHBAZ. 2007b. New and noteworthy species of *Boechera* (Brassicaceae) III: additional sexual diploids and apomictic hybrids. Harvard Papers of Botany 12: 235–257.

LEGENDS SUPPLEMENTARY MATERIAL

Supplementary figure 1 Geographic distribution of accessions used in this study

Supplementary table 1 Accession list showing internal accession number, ITS type, trnLF haplotype as well as origin of the herbarium specimen, taxon information and geographical location.

Supplementary Table 2 ITS type gene bank accession numbers and accessions sharing the ITS-type as well as cpDNA haplotype identified and the ecoregion to which the accession was assigned. Colour codes indicate if the ITS type occurred once (yellow), if it was species specific (orange) or if it was shared by several taxa (red). Bold letters indicate the presence of ambiguous codes in the sequence.

Supplementary table 3 Annotated Alignment including all ITS types recovered in this study and in Koch et al. 2003.

Supplementary table 4 ITS Alignment reduced to ITS types without ambiguous codes other than N and originating from accessions which were most likely not apomictic according to the species determination.

Supplementary table 5 Alignment of the intron sequences from the intron of At2g25920 neighbouring the ELF3 gene.

Supplementary table 6 Alignment of the intron sequences from the intron of At3g18900 neighbouring the VRN1 gene.

Supplementary table 7 Table listing material used for the single copy gene intron analysis together with ITS types and chloroplast haplotypes.

2.3 Eastern versus western North American *Boechera*

Manuscript submitted to TAXON;

supplementary data are included on the DVD in the back of the dissertation.

Running title:
Systematics of eastern North American *Boechera*

Boechera or not? Phylogeny and phylogeography of eastern North American *Boechera* species (Brassicaceae)

Christiane Kiefer, Christoph Dobeš, and Marcus A. Koch[#]

Heidelberg Institute of Plant Science, Biodiversity and Plant Systematics, Heidelberg University, Heidelberg, Germany

Correspondence:
[#] Marcus A. Koch. Heidelberg Institute of Plant Science, Department of Biodiversity and Plant Systematics, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany. Fax: +49-(0)-6221-54-5508; E-mail: mkoch@hip.uni-heidelberg.de

[1]Manuscript received _____; revision accepted _____.

Acknowledgements
We are grateful to the curators of the herbaria GH, MO and DAO for providing leaf material, and Ihsan Al-Shehbaz for continuous advice and support during our studies on *Boechera* . Funding from DFG to M.A.K. (German Research Foundation) was greatly acknowledged.

Abstract

The North American genus *Boechera* comprises according to current taxonomic classification 110 species. Most of them are centered in western North America, with some species extending their distribution range from the west into eastern North America and seven species exclusively occur in eastern North America. Past phylogenetic studies included at most two of the eastern species and placed them as sister to two representatives of western North American *Boechera*. Our studies based on chloroplast DNA markers (*trn*L intron, *trn*LF IGS) and nuclear ribosomal DNA ITS1 and ITS2 show that eastern North American *Boechera* represent two evolutionary lineages within the genus *Boechera* (cpDNA) or among the tribe Boechereae (nrDNA). We included also ancestral *Polyctenium* and other Boechereae like *Sandbergia* and *Borodinia*, for which close relationships to *Boechera* are assumed. Based on our data we suggest to retain eastern North American *Boechera* within *Boechera* rather than transferring them into a new genus. Some phylogeographic conclusions are also drawn like a possible migration scenario of *Boechera* into eastern North America. The Siberian species *Boechera falcata* as well as the Siberian genus *Borodinia* were shown to be placed within *Boechera*.

KEYWORDS: *Boechera*, *Borodinia*, Brassicaceae, cpDNA, ITS, phylogeny, phylogeography, *Sandbergia*, *trn*LF

INTRODUCTION

Boechera A. Löve & D. Löve is a North American genus from the Brassicaceae (mustard family) and is currently subject to a broad range of studies such as apomixis research (Sharbel & al. 2004, Voigt & al. 2007), phylogeography (Dobeš & al., 2004a), drought tolerance (Knight & al., 2006), host-fungi interactions (Roy, 2001), and response to herbivory (Siemens & al., 2003). Furthermore, genome and chromosome evolution were studied in great detail (Kantama & al., 2007). Population studies were carried out using the sexual species *B. stricta* (Song & al., 2006) and crossing experiments revealed the high potential for hybridisation within *Boechera* (Schranz & al., 2005). Based on recent molecular studies and analysis of morphological data (Al-Shehbaz & al. 2006) *Boechera* belongs to the tribe Boechereae together with six other genera [*Anelsonia, Cusickiella*, *Nevada*, *Phoenicaulis, Polyctenium*, *Sandbergia* (= *Halimolobus perplexa*, Al-Shehbaz, 2007)]. The Boechereae are almost exclusively North American, have a base chromosome number of x=7 (Dobeš & al., 2006, Warwick & al., 2006) and are characterized with few exceptions by branched trichomes and entire leaves. Most of the members of the tribe are perennials with a well-developed basal rosette (Al-Shehbaz & al., 2006). *Boechera* has its highest species diversity in western North America. However, seven of the 110 currently accepted species (Windham & Al-Shehbaz, 2006, 2007a, 2007b) are restricted to eastern North America (*Boechera laevigata*, *Boechera missouriensis*, *Boechera perstellata*, *Boechera burkii*, *Boechera shortii*, *Boechera serotina*, *Boechera canadensis*; a summary of the distribution ranges of eastern North American *Boechera* species and their habitats and chromosome numbers is given in Table 1). Eastern and western North American *Boechera* differ from each other in life form, habitat preferences, as well as morphology. The eastern species are biennials of forest habitats and have dentate obovate to broadly oblanceolate leaves up to 16 cm long. In contrast, western *Boechera*s are mainly perennials of open habitats and have entire, much smaller leaves (Rollins, 1993).

A phylogenetic study based on the plastid *ndh*F placed eastern North American *B. laevigata* and *B. shortii* as sister to *Boechera platysperma* and *Anelsonia* (Beilstein & al., 2006). This sister group relationship of *B. laevigata* and *B. platypserma* was also shown in a study based on several plastidic and nuclear markers which in addition included *Boechera stricta* which was phylogenetically closest to *B. platysperma* (Bailey & al., 2006). In both studies *Polyctenium* was sister to all other members of tribe Boechereae, followed by *Cusickiella* and *Nevada* as next closest sisters. However, the phylogenetic position of *Cusickiella* seems to depend strongly on the applied marker system which is obvious from an earlier phylogenetic study where three different marker systems were applied (*trn*LF region, nrDNA ITS, *pistillata* intron 1; Bailey & al., 2002). No other eastern North American *Boechera* species except *B. laevigata* and *B. shortii* were included in any of these

phylogenetic reconstructions so no information about their relatedness is available so far. Furthermore *Boechera falcata*, the only Siberian *Boechera* described so far, as well as the Siberian genus *Borodinia* for which a close relationship to *Boechera* was assumed from morphological data, have not been included into a molecular phylogenetic study yet. At the generic level past studies of the chloroplast DNA marker *trn*LF revealed three major evolutionary lineages within *Boechera* (Dobeš & al., 2004) and were complemented by the analysis of nuclear encoded loci (Koch & al., 2003, Dobeš & al., 2004a). These studies included only *B. stricta* (sexual, diploid species), *B. holboellii* sensu Rollins and *B. divaricarpa* sensu Rollins (Rollins, 1993) but none of the eastern North American relatives.

Table 1 Distribution ranges, habitats and chromosome numbers for eastern North American *Boechera*.

species	distribution	habitat	Base chromosome number
Boechera laevigata	southern Quebec southwards into Georgia and Alabama, west into Oklahoma and Kansas, and north into Minnesota	in wooded areas and on cliffs, sandhills, bluffs, and limestone ledges	n=x=7
Boechera missouriensis	from Maine and Vermont south into Georgia, west through Oklahoma, and north into Wisconsin	in wooded areas and on cliffs, sandhills, bluffs, and limestone ledges	n=x=7
Boechera canadensis	from Canada's southern Quebec and Ontario south into Florida, westwards into Texas, and north into Nebraska and Minnesota	Riparian habitats	n=x=7
Boechera burkii	Maryland, Pennsylvania., Tennessee, Virginia and West Virginia	in rocky areas, wooded slopes, and on river banks	Not available
Boechera serotina	Virginia and West Virginia	on shale barrens and wooded slopes of crumbling shale	2n=2x=14
Boechera shortii	from New York south into Alabama, west into Kansas, and north into South Dakota and Minnesota	in rich woods, on stream banks, lake shores, steep wooded slopes, limestone bluffs and cliffs, river flood plains, and shaded bottomlands	2n=2x=12 (counted in *Arabis dentata* (Smith 1938)= *Boechera shortii*)
Boechera perstellata	Kentucky and Tennessee	hillsides and calcareous bluffs	2n=2x=14

Palynological studies have shown that during quaternary glaciation cycles forest habitats underwent relocations in eastern North America as tree species retreated into refuge areas further south avoiding the colder temperatures (Jackson & al. 2000), and one might assume that the *Boechera*s followed migration of these habitat types. However, some tree species as for example *Quercus rubra* remained further North closer to the ice shield during the LGM (Magni & al. 2005). Actually, nothing is known about the phylogeographic history of eastern North American *Boechera* species. For *B. stricta*, whose distribution range extends into north eastern America chloroplast DNA data indicated a potential refuge area near the Great Lakes region (Dobeš & al. 2004a). However, this was not true for *B. holboellii* varieties investigated in the same study. A well known phylogeographic pattern in eastern North America is the Atlantic and Gulf Coastal Plain disjunction to the Great Lakes region. *Prunus* (Shaw & Small 2005) as well as *Trillium grandiflorum*, a herbaceous woodland species (Griffin & Barrett 2004), seem to follow this pattern among many others.

Inhere, first we aim to reconstruct the phylogenetic relationships between the eastern North American *Boechera* species and those centred in western North America as well as the most of the remaining Boechereae based on cpDNA (*trn*L intron and *trn*L-F intergenic spacer) and nrDNA (ITS1, ITS2). Here we included also *Borodinia* and *Boechera falcata*, for their phylogenetic positions within Boechereae are still unclear.. Second, we would like to verify the morphology-based generic assignment of eastern North American *Boechera* using the same data. Third, based on a phylogeographic analysis of the chloroplast DNA we will try to identify potential refuge areas of eastern North American *Boechera* species.

MATERIALS AND METHODS

Plant material.—Leaf material was obtained from herbarium specimens from GH, MO and DAO. Corresponding accession details are listed in Table S1 (Online supplementary material). From *B. burkii* we obtained no and from *B. serotina,* we obtained only limited material, and only some molecular DNA sequence data from herbarium material of *B. serotina* were obtained successfully.

DNA extraction.—Total DNA was obtained from a 5x5 mm^2 piece of dried leaf tissue from single individuals. Extraction followed the CTAB method of Doyle & Doyle (1987), but some modifications were applied, involving grinding of only a 5x5 cm^2 piece of of dry leaf tissue in 2ml tubes using a Retsch swing mill (MM 200), addition of two units of ribonuclease (RNAse A) to the resolved DNA, and washing of the DNA pellet twice with 70% ethanol. DNA was finally dissolved in 50-70µl low TE-buffer or low Tris-buffer for long-term storage.

PCR conditions.—PCR reactions were performed in a volume of 25 µl containing 1x GoTaq buffer (Promega), 2 mM $MgCl_2$, 5 pmol of each primer, 5 nmol dNTPs (1.25 nmol of each dNTP) and one unit Taq DNA polymerase (GoTaq, Promega), and variable concentrations of template (50 to 400 ng) using a PTC-200 thermal cycler (MJ-Research). Thermal cycling started with a denaturation step at 95°C lasting three min; followed by 30 cycles each comprising 30 s denaturation at 95°C, 30 s annealing at 48°C and elongation at 72°C. Amplification ended with an elongation phase at 72°C lasting 10 min, and a final hold at 4°C.
The *trn*L intron was amplified using the forward primer 5‘-CGA AAT CGG TAG ACG CTA CG-3‘ and the reverse primer 5‘-GGG GAT AGA GGG ACT TGA AC-3‘ (primer c and d according to Taberlet & al., 1991), which anneals in the first and second exon of the *trn*L gene, respectively. Sequences comprised the complete intron and the second exon of the *trn*L gene. For amplification of the *trn*L-F intergenic spacer (IGS) primers 5‘-GGT TCA AGT CCC TCT ATC CC-3‘ (primer e according to Taberlet & al., 1991) and 5‘-GAT TTT CAG TCC TCT GCT CTA C-3’ (designed in Dobeš & al., 2004a) annealing in the second exon of the *trn*L gene and the *trn*F gene, respectively, were used. Amplified sequences included the complete IGS and the first 18 bases of the *trn*F gene. PCR products were checked for length and concentrations on agarose gels (1% agarose in TAE). PCR products were all purified with the NucleoFast Kit (Macherey & Nagel, Germany).
ITS1 and ITS2 were amplified together as described in Dobeš & al. (2004b). Sequences comprised the complete ITS1 and ITS2 as well as the 5.8s rRNA gene. PCR products were checked on agarose gels (1% agarose in TAE). PCR products were all purified with the NucleoFastKit (Macherey & Nagel, Germany).

Cycle Sequencing.—Cycle sequencing was done with the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences) using the PCR primers for the cycle sequencing reaction. Samples were resolved in 10 µl Loading Solution and then run on a MegaBace 500 Sequencer.

Alignments and chloroplast haplotype/ITS type definition.—After sequencing, the forward and reverse sequences were aligned, edited by hand and trimmed to a common length. The *trn*L intron and the *trn*L-F IGS region were assembled into one sequence (*trn*LF region) and missing bases at the joining 3’- and 5’-primer ends were substituted with Ns. Duplicate sequences were removed and cpDNA and ITS haplotypes were defined based on single nucleotide and indel polymorphisms. Haplotypes designation introduced by Dobeš & al. (2004) and Schranz & al. (2005) were assigned to those haplotypes already published by these authors and found again in this study. The alignments used for the phylogenetic analysis were made manually using the program GenDoc (Nicholas & Nicholas, 1997) according to the

alignment published by Dobeš & al. (2004) including one to three *trn*F pseudogenes at the 3´-end of the *trn*LF intergenic spacer. ITS types were defined by the same method as described for *trn*L-F haplotypes.

***Trn*LF Phylogenetic reconstruction.—**For reconstruction of phylogenetic relationships of taxa belonging to the Boechereae sequences of *Sandbergia* (EU850744, EU850745), *Cusickiella* (EU850746, EU850747; EU850748, EU850749; EU850750, EU850751; EU850754, EU850755; EU850752, EU850753) and *Polyctenium* (EU850756, EU850757; EU850758, EU850759; EU850760, EU850761; EU850762, EU850763; EU850764, EU850765; EU850766, EU850767; EU850768, EU850769) as well as several sequences representing members of the Halimolobeae (AF307555, AF307549, AF307551, AF307543, AF307545, AF307538, AF307530, AF307531) were added to the alignment used for the network analysis (see below). *Capsella bursa-pastoris* (AY122454, DQ310514) was used as outgroup. The region of the *trn*L second exon was omitted from the analysis (positions 530-610) because many missing or ambiguous data were present in this region. Tree reconstructions were done by using TREECON for Windows 1.3b (Van de Peer & De Waachter, 1994). A neighbour-joining (NJ) analysis (Kimura-2-parameter distance; transition/transversion ratio calculated and set automatically from the data; gaps excluded) was run excluding the pseudogene region and its further 3´-region (bp. 802-1214) because of uncertain homology of the pseudogenes found in different genera (see haplotype definition for more information on pseudogenes). A bootstrap analysis running 1000 replicates was performed with TREECON using the same distance measurements.
In order to estimate the potential influence of the various gaps on tree topology we performed the same analysis running TREECON with insertion/deletions taken into account.

***Trn*LF network reconstructions.**—For network analysis the most common haplotypes (AY257725, AY257736, AY257718, AY257692, AY257710, AY257694, AY257712, AY257697, AY257703, AY257761, AY257764, AY257775, DQ012052, AY257778) identified by Dobeš & al. (2004) were added to the alignment to infer the relationship of haplotypes of eastern North American *Boechera* and previously identified haplotypes. Network reconstruction was done by running the program TCS1.21 (Clement & al. 2000), in which the gaps were coded and the connection limit was set to 95%. The analysis was run excluding the *trn*F pseudogenes (see haplotype definition for more information on pseudogenes).

ITS phylogenetic tree reconstructions.—Phylogenetic reconstructions based on ITS1 and ITS2 were done by running parsimony analyses as implemented in PAUP4.0beta (Swofford, 2002). The maximum of trees retained was limited to

10,000. For the heuristic search sequences were added randomly in 1000 replicates during which 10 trees were saved in each replicate. Chuckscore was set to one. Gaps were treated as missing. TBR was used as branch-swapping algorithm. Starting trees were obtained via stepwise addition. Six characters were excluded from the analysis (505-510). All characters had equal weight.
For phylogenetic reconstructions based on the ITS dataset alone the alignment was supplemented with sequences of closely related genera obtained from Genebank to determine the relationship of eastern North American *Boechera* species to representatives from western North America and to other genera of the tribe *Boechereae* (DQ452606, AF146515, AF146514 = *Cusickiella douglasii*, DQ452059 = *Anelsonia eurycarpa*, DQ452061 = *Nevada holmgrenii*, DQ452066 = *Cusickiella quadricostata* and AJ232927, AJ232926 = *Halimolobus perplexa* var. *lemhiensis* (hereafter *Sandbergia*)). Three *Capsella* species served as outgroup (*Capsella rubella* AJ232913, *Capsella bursa-pastoris* AF137570, *Capsella grandiflora* AM905718). In order to resolve relationships among the Boechereae and Halimolobeae *Halimolobos* (AF307634, AF307635), *Mancoa* (AF307631), *Sphaerocardamum* (AF307611, AF307612) and *Pennellia* (AF307627, AF307629), *Arabis tricornuta* (AF307628) were added to the analysis.
Bootstrap analysis was also run in PAUP4.0beta with the same settings as the parsimony analysis in 1000 replicates.

Geographical analysis.—All information from the herbarium vouchers was entered into the database BioOffice (Biogis Consulting, Version 2.0.4). Missing geographical coordinates were added according to the descriptions on the herbarium vouchers. Using BioOffice haplotypes were plotted on North America maps included in the ArcView Package Version 8.

RESULTS

Alignments.—The length of the alignment of the combined *trn*LF region used for intrageneric analysis in *Boechera* was 975 bp (Table S2, online suppl. material; *trn*L intron: position 1 to 506; *trn*L second exon: position 507 to 556; the *trn*L-F IGS: position 557 to 960, the partial sequence of the *trn*F gene from 961 to 975) in the alignment for the network analysis. In the *trn*L-F IGS one to three pseudogenic copies of *trn*F were present as in all other *Boechera* species and closely related genera like *Halimolobus*. The pseudogene region (alignment position 626-933) was excluded from all further analyses except for haplotype definition because addition and deletion of copies does not coincide with the phylogenetic signal of regions other than the pseudogenic one (cf. Dobeš & al., 2007).

The *trn*LF alignment for the phylogenetic analysis including the other genera from the Boechereae and *Capsella* species as outgroup had a total length of 1214 bp. The alignment is given in Table S4.

The ITS alignment had a total length of 693 bp (Table S3, online suppl. material; partial sequence of the 18s rRNA gene: position 1 to 23; ITS1: position 24 to 296; 5.8s rRNA gene: positions 297 to 469; ITS2: position 470 to 693).

***Trn*LF phylogenetic reconstruction.**—Phylogenetic reconstruction based on the *trn*LF dataset is shown in Fig. 1. The representatives of the Halimolobeae included in this study were sister to the Boechereae with moderate bootstrap support. The

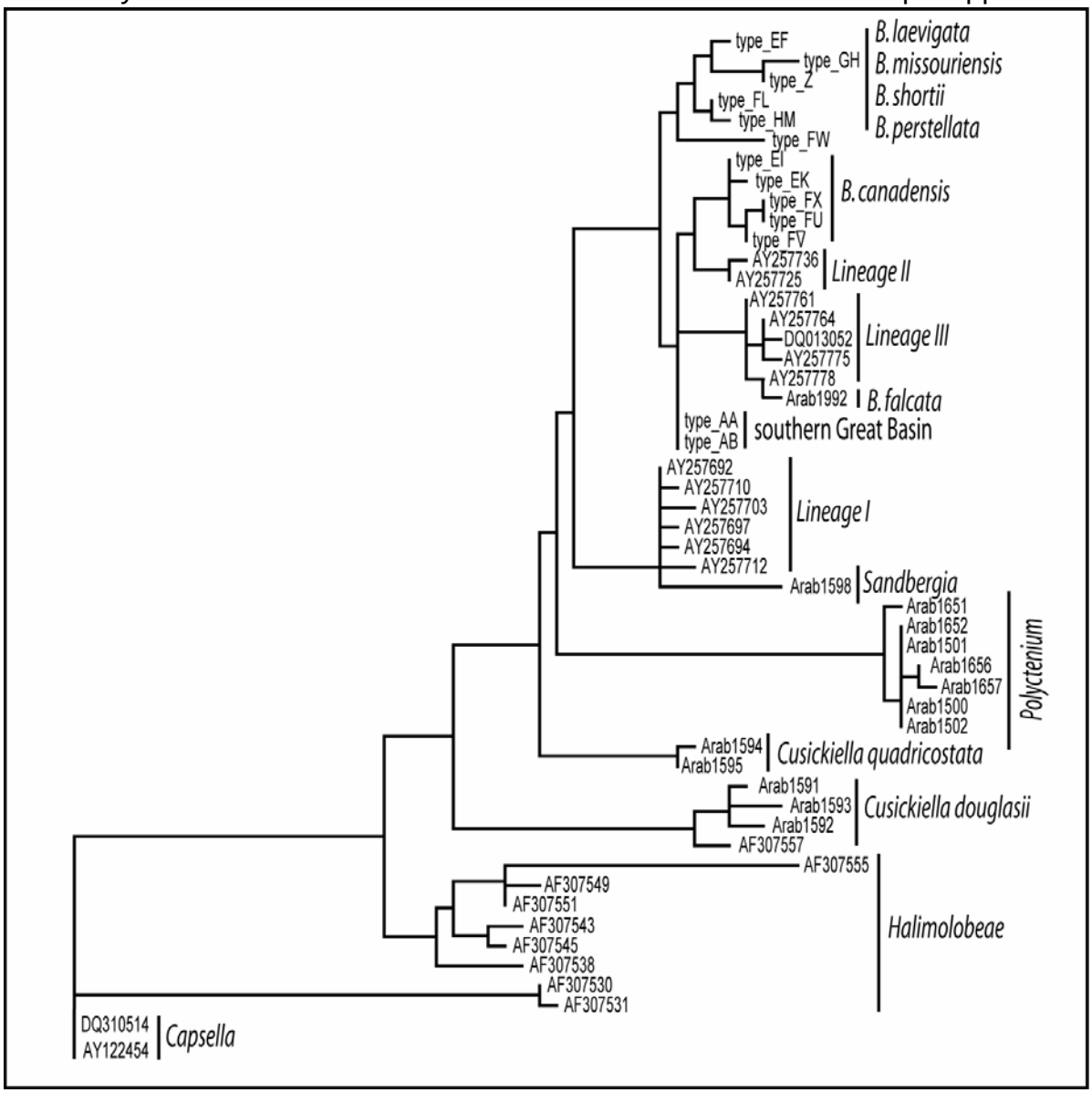

Figure 1 Neighbour-joining tree based on cpDNA marker system *trn*LF; for the outgroup (*Capsella*) and the Halimolobeae (Bailey & al. 2002) genebank accession numbers are indicated, for the Boechereae either the haplotype name (*Boechera* species except *Boechera falcata*) or the accession number of the individual in our study is given; for information on the individuals carrying those haplotypes refer to supplementary Table 1. *Arabis tricornuta* (AF307555= has been proposed to be transferred to the genus *Pennellia* (Al-Shehbaz 2003).

genus *Cusickiella* constituted two lineages representing C. *douglasii* and *C. quadricostata*, respectively, at the base of the Boechereae and was paraphyletic to a clade comprising *Polyctenium fremontii*, *Sandbergia* and *Boechera*. *Cusickiella* was not grouped into a monophyletic lineage and *Polyctenium* was placed ancestral to all remaining taxa of the Boechereae. *Sandbergia* grouped together with *Boechera* lineage I haplotypes (sensu Dobeš & al. 2004a). *Boechera* lineage I and *Sandbergia* were sister to a clade comprising *Boechera* lineages II and III, all eastern North American *Boechera* species as well as two haplotypes (AA, AB) found exclusively in the southern Great Basin. Siberian *Boechera falcata* was most closely related to haplotype CI (AY257778), a wide spread haplotype from cpDNA evolutionary lineage III.
Analysis running under the OPTION “insertion/deletions” resulted in an almost identical topology (data not shown) with two notable changes: 1) *Sphaerocardamum* and remaining Halimolobeae are monophyletic with weak bootstrap support (BS 53%) and sister to Boechereae. 2) *Sandbergia* is ancestral to all other *Boechera* species with weak bootstrap support as well (BS 41%), a finding consistent with ITS data as described below.

ITS Phylogenetic reconstructions.—In the ITS dataset 687 characters were included in the analysis with 461 characters found to be constant, 69 parsimony-uninformative characters and 157 parsimony-informative characters. We excluded characters 505-510, which could be hardly aligned (Supplementary Table 3.) The phylogenetic reconstruction recovered a strict consensus tree based on 1,000 shortest trees with a good resolution among terminal accessions corresponding to species or groups of species, but in contrast with a very poorly supported relationships among them (Fig. 2) (CI=0.76, RI=0.92, RC=0.70, tree length=370).
Cusickiella was sister to the Halimolobeae and Boechereae (bootstrap support (BS) = 40%). The Halimolobeae and Boechereae were also found to have a sister group relationship. However this was only weakly supported (BS 38%).

Within the Boechereae several lineages were identified: three lineages were found to arise from a basal polytomy. The first lineage included *Anelsonia*, the second *Sandbergia* and *Polyctenium*, and the third *Borodinia*, *Nevada* and eastern and western North American *Boechera*. *Borodinia* was on a polytomy with a lineage comprising *B. perstellata*, *B. shortii*, *B. missouriensis*, *B. laevigata* and *B. serotina* and a lineage comprising *B. falcata*, western North American *Boechera* and *B. canadensis*. The splits differentiating major groups had low to no bootstrap support.
In general it is noteworthy that among the eastern North American *Boechera* species the ITS is not species-specific. This is also true for most western North American *Boechera* species (Kiefer & Koch, unpublished).

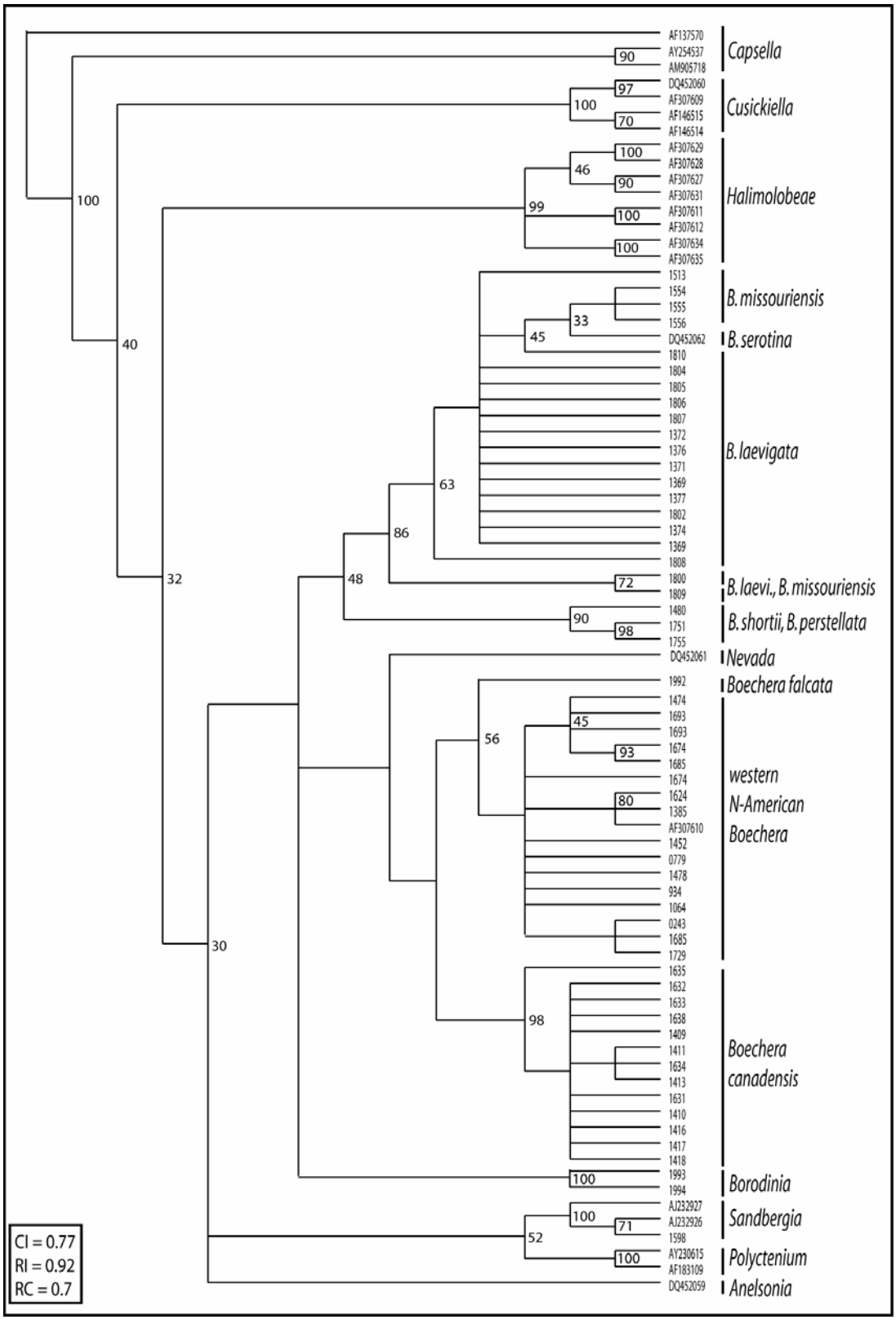

Figure 2 Parsimony analysis of nrDNA data; strict consensus tree of 10,000 shortest trees. *Capsella* was used as outgroup. For the Halimolobeae, *Cusickiella*, *Polyctenium*, *Anelsonia* and two *Sandbergia* accessions genebank accession numbers are given; for *Boechera* species either a genebank accession number is given if it was analysed in a previous study or the internal accession number from supplementary Table 1 is given.

***Trn*LF network.—**The network analysis revealed six different lineages. It recovered the previously described lineages I, II and III (Dobeš & al. 2004a) and two additional lineages described herein as lineages IV and V (Fig. 3) which contained the eastern North American *Boechera* species and a lineage containing only one haplotype. Haplotypes AA and AB (combined here to one single haplotype because they differ only in one mutation in the excluded pseudogene region) were found to be in the centre of the network. Lineage IV contained five haplotypes that were exclusive to *B. canadensis* with the most internal haplotype FX being connected to AA/AB by seven mutational steps. Lineage V was connected to the central haplotypes AA/AB by one mutational step (haplotype EF). Haplotype FL was derived from haplotype EF by one mutational steps and haplotype HM was derived from haplotype FL by one mutational step. Haplotype Z (three mutational steps) and GH (four mutational steps) were also derived from haplotype EF. Haplotype EF was private to *B. laevigata* with one exception where it was also found in a *B. missouriensis* individual. Haplotype FL was found in *B. missouriensis*, *B. perstellata*, and one individual of *B. laevigata*. Haplotype GH and Z were found in *B. shortii*. Haplotype Z was also found in one *B. divaricarpa* accession [taxonomy according to Rollins (1993)], a hybrid species of multiple origin (see Windham & Al-Shehbaz, 2007b).

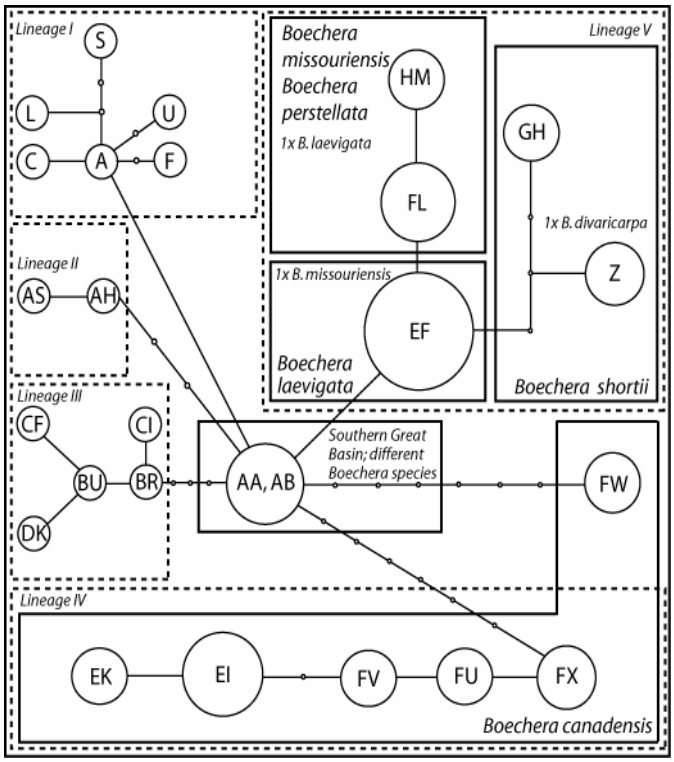

Figure 3 Network analysis of *trn*LF data to reconstruct phylo-geographic history of eastern North American *Boechera* species. These species are found in three distinct lineages which like the haplotype lineages typical for western North American *Boechera* species originate from Haplotype AA/AB. Haplotype names are given as letter codes; individuals carrying these haplotypes can be found in supplementary Table 1. Boxes with dashed lines encircle lineages, boxes with solid lines indicate haplotypes specific to a single species or a group of closely related species.

Phylogeographic analysis.—For phylogeographic analysis, the geographic distribution of haplotypes was plotted onto a map (Fig. 4a, b). For *B. canadensis,* haplotype EI was found north of the last maximum glaciation in New York and north of Lake Erie in Ontario and southern Quebec. Only one accession in Iowa was found south, but still close, to the maximum extent of the glaciers during the last

glaciation. Haplotype EK, derived from EI, was found at the western edge of the distribution range of EI also in Iowa. Haplotypes FU and FX were found in Arkansas and Georgia, and haplotype FV was found in Mississippi. Haplotype FW was found east of the Appalachian Mountains in North Carolina.

Boechera laevigata haplotype EF was found throughout the sampled area. The northern range of EF extended along lakes Erie and Ontario, as well as into New York State and southern Quebec. The southern range of EF included Virginia, North Carolina, and Tennessee. In North Carolina bordering Tennessee, the haplotype EF was also shared by one *B. missouriensis* accession. Haplotype FL was distributed south of the last glaciation co-occurring with EF. It was found in Tennessee, South Carolina, North Carolina, and Oklahoma. In Tennessee FL was also shared by *B. perstellata* and one *B. laevigata* accession. *Boechera shortii* haplotype GH was found in Missouri, and its haplotype Z was found north of Erie Lake. Haplotype Z, shared by *B. divaricarpa,* was found in Minnesota.

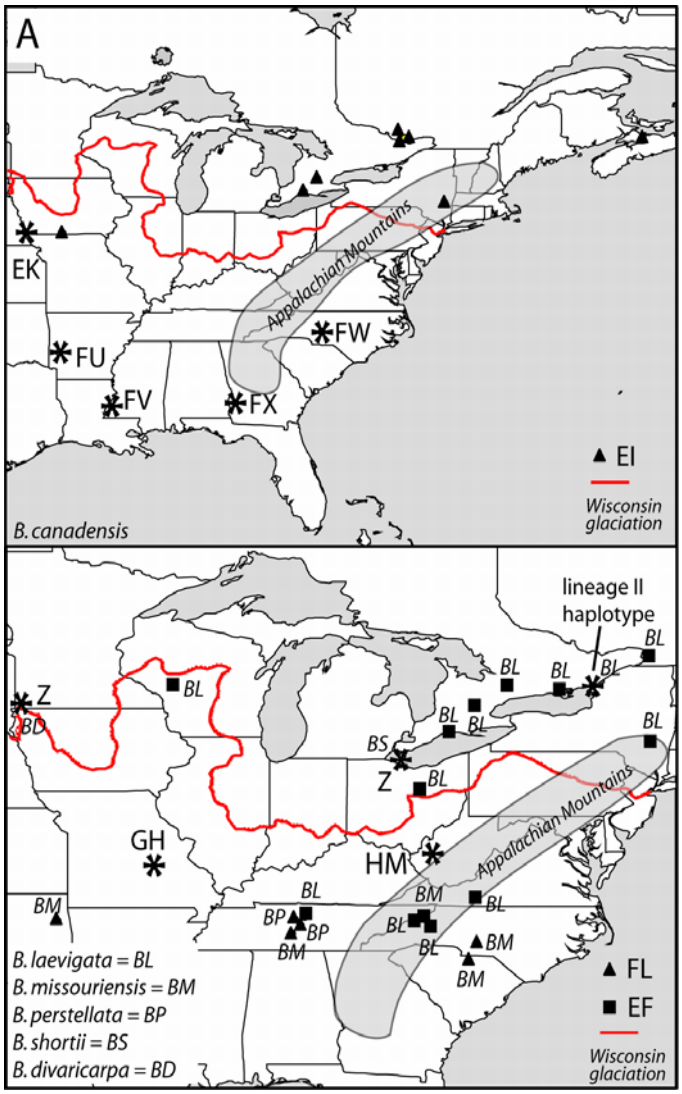

Figure 4 Distribution ranges of haplotypes found in eastern North American *Boechera*. The extension of the last glaciation is shown as a red line, the Appalachian Mountains are shown as shaded area, species names in figure (B) are coded with two letters; different symbols represent different haplotypes whose name is given in the figures; asterisks indicate singleton haplotypes. (A) Lineage IV; haploytypes in lineage IV were all found in *B. canadensis* accessions. Haplotypes found in the south are connected to the centre of the network shown in figure 3; north of the extension of the last glaciation only haplotype EI is found. (B) Lineage V; north of the extension of the last glaciation only haplotypes EF and Z were found. South of the extension of the last glaciation haplotype EF occurred together with haplotype FL and two derived haplotypes.

DISCUSSION

Phylogenetic Reconstruction.— Eastern and western North American *Boechera* species differ significantly in their life history, habitat preferences and morphology. The question is if this differentiation can also be recognized in the divergence of marker sequences and how Eastern North American *Boechera* are related to other genera within the Boechereae. Our study based on cpDNA placed *B. laevigata*, *B. missouriensis*, *B. shortii* and *B. perstellata* inside *Boechera* as sister to *Boechera* chloroplast lineages II and III and *B. canadensis*. In the nrDNA dataset on the other hand *B. laevigata*, *B. missouriensis*, *B. perstellata*, *B. shortii* and *B. serotina* were sister to ITS-lineages representing western North American *Boechera*. In a family-wide phylogeny *B. laevigata* was found to be sister to *B. platysperma* and *B. stricta* (Bailey & al. 2006). A phylogenetic study based on *ndh*F placed *B. laevigata* and *B. shortii* as sister group to *B. platysperma* and *Anelsonia* (Beilstein & al. 2006). Both results are in agreement with our result obtained for the ITS analysis.

Boechera canadensis was not included into previous phylogenetic studies. Inhere *B. canadensis* was carrying haplotypes nested within *Boechera*, sister to *Boechera* chloroplast lineage II. Incongruencies between phylogenies based on different marker systems can occur due to different reasons. Within *Boechera* it seems likely that we have to explain distribution of chloroplast DNA variation among species as a result of the distribution of ancestral genetic variation predating separation of the various lineages carrying these haplotypes today. This phenomenon has been described in detail not only for *Boechera* (Dobeš & al, 2004), but also for the genus *Arabidopsis* (Koch & Matschinger, 2007).

The Siberian *B. falcata* has also not been included in phylogenetic analyses so far. In our analyses it was placed with western North American *Boechera* species for cpDNA as well as nrDNA ITS. The cpDNA haplotype found in *B. falcata* was found to be related to haplotype CI (AY257778) found in evolutionary lineage III. This haplotype is among the most widespread haplotypes within *Boechera* as was shown in previous phylogeographic studies (Dobeš & al., 2004a). It also extends into Alaska which might hint on a migration into Siberia via the Bering Strait. The ITS sequence from *B. falcata* was sister to ITS sequences found in western North American *Boechera* which suggests an early split of *B. falcata* and western North American *Boechera* species.

Resolution between *Boechera* and other members of the Boechereae was low and our analyses were partially in conflict with the analyses performed in previous studies (Bailey & al., 2002; Bailey & al., 2006; Beilstein & al., 2006). However, we can show that the Siberian genus *Borodinia* groups together with *Boechera*. Hereby, the placement of *Borodinia* within the Boechereae or even *Boechera* based on morphological data can be confirmed by molecular data. Furthermore, we found weak evidence that *Sandbergia* is basal to *Boechera* and *Borodinia*. The placement

of *Cusickiella* was dependent on which marker system was applied. The cpDNA dataset suggested to place *Cusickiella* within tribe Boechereae while the analysis of nrDNA data places it even outside the Halimolobeae. This discrepancy in the placement of *Cusickiella* was also observed in a study of the Halimolobeae where *Cusickiella* and *B. stricta* (former *Arabis drummondii*) were used as representatives of the Boechereae (Bailey & al. 2002). In a family wide phylogeny based on several nuclear and plastidic marker systems *Cusickiella* was resolved within the Boechereae and *Polyctenium* was sister to *Cusickiella* and *Boechera* (Bailey & al., 2006). This is in conflict with both of our phylogenetic analyses. However, bootstrap support for the relationships among the Boechereae was almost low in our datasets and it was also low in the family wide phylogeny. At this point we can conclude only that *Cusickiella* is paraphyletic and *Polyctenium* has been considered for any future taxonomical changes within this group. The family wide phylogeny was already based on ten different marker systems (five from the plastid including *trn*LF, five from the nuclear genome including ITS), and might also reflect the general conflict between nuclear and plastid derived phylogenetic hypothesis. So it is unlikely that the addition of more sequence data from the various genomes would lead to a higher bootstrap support or better resolution for the relationships among the Boechereae.

Phylogeographic Analysis based on the *trn*LF Network.—Our sampling of eastern North American *Boechera* species covered their entire distribution ranges, and so we could draw some conclusions on their phylogeographic history. Phylogeographic studies of eastern North America have revealed a multitude of different patterns in the past (Soltis & al., 2005). Many of the patterns are shaped by the Appalachian mountains that in some cases acted as barriers to gene flow (Griffin & al., 2004; Godbout & al., 2005). In the case of *B. laevigata* (haplotype EF only) and *B. missouriensis* (haplotypes FL and EF only) this mountain chain seems to have had little effect on haplotype distribution and genetic differentiation as the same haplotypes are found east end west of the ridge of the Appalachian mountain chain. In *B. canadensis* sampling density is not sufficient to draw any conclusions on that particular pattern.

A common pattern of phylogeographic analyses is northern purity versus southern richness as an effect of the quaternary ice ages (southern refugia and recolonization of northern areas) (Hewitt, 2000). For *B. canadensis* the basal haplotypes were all found south of the last glaciation in southeastern North America and Iowa while north of the last glaciation one single derived haplotype was found. The second chloroplast DNA lineage for eastern North American *Boechera* was remarkably poor in haplotypes considering the fact that it contains four species and it was not possible to draw any conclusions on the phylogeographic history of this lineage. The extreme haplotype poverty in this lineage might hint at recent differentiation of those taxa. In

the past *B. laevigata* and *B. missouriensis* were sometimes even treated as varieties of the same species (*Arabis laevigata* (Muhl. ex. Willd.) DC. var. missouriensis (Greene) H.E. Ahles; Ahles, 1964). This view is also supported by some haplotype sharing between *B. missouriensis* and *B. laevigata*. Haplotype sharing of *B. laevigata* and *B. missouriensis* occurred in Tennessee in the Interior Plateau ecoregion adjacent to the southwestern Appalachian Mountains. Rollins (1993) already noted that "… in North Carolina in particular, the two species appear to come close together morphologically …". In fact this might indicate hybridisation of the two species in this region or regions close by (like in Tennessee) as is also indicated by the chloroplast data.
B. laevigata and *B. missouriensis* ecologically favour forest habitats. From the descriptions on herbarium specimens and from the placement of accessions on the used ecoregion map (Olson & al., 2001) the analysed individuals were all collected in temperate broad leaf and mixed forest in particular in connection with various *Quercus* species. During the last glacial maximum *Quercus* was abundant in the Mississippi valley and Florida but it also occurred in patches in between (Jackson & al. 2000). If *B. laevigata* and *B. missouriensis* had to retreat into refugia during the last glacial maximum they may have followed *Quercus* in one of the scattered populations and may have hybridized there as a consequence of close spatial contact.

B. shortii is only represented by two individuals in the study. Its haplotypes are derived from EF and form a lineage with many missing haplotypes. From two individuals no phylogeographic history can be inferred but it is still noteworthy that haplotype Z found in *B. shortii* is also shared by one triploid apomictic hybrid that previously had been identified as *B. divaricarpa.* According to the currently accepted description this species is restricted to Washington, Idaho, and Montana, south to Wyoming, Utah, Nevada, and California (Windham & Al-Shehbaz, 2007b). However, in the past *B. divaricarpa* was used as a "waste basket species" (Windham & Al-Shehbaz, 2007) and many hybrids of different parentage were determined as *B. divaricarpa*. The individual described as B. divaricarpa should be revisited for species identity as typically *B. divaricarpa* shares haplotypes from lineages I, II or III (Dobeš & al., 2004).

Haplotype AA/AB is central to the cpDNA network and connects the five identified lineages to each other. Today it is only found in the southern Great Basin. This raises the question where eastern North American *Boechera* species evolved and how they arrived in their current distribution range. Most *Boechera* species originated in western North America, an area from which lineages IV and V reached the eastern part of the continent, as evidenced by the distribution range of central haplotype AA/AB from which lineages IV and V are derived. From our data the

following hypothesis about the evolutionary relationships between western and eastern North American haplotypes may be developed: Haplotype EF probably originated in western North America prior to glaciation or during interglacial periods and later migrated towards the East. Haplotypes Z and GH also differentiated from EF probably when lineages IV and V were not fully differentiated from lineages I, II, and III and the distribution rages of species from all lineages overlapped. During that time, haplotype Z was captured by species that normally carry haplotypes from lineages I, II or III like for example "*B. divaricarpa"* or its ancestral species. Haplotypes EF and Z became extinct in the West because of repeated glacial cycles and survived only in eastern North America. Haplotype Z was also diminished and many of the intermediate haplotypes leading from EF to Z were lost.

SUMMARY

Our phylogenetic reconstructions based on chloroplast DNA and nrDNA were not congruent. This can most likely be explained by the different evolutionary histories of the employed marker systems. Support for ancestral splits between lineages in the phylogenetic reconstruction based on nrDNA data was low while more shallow splits had a good support. Therefore we can conclude that *B. laevigata*, *B. missouriensis*, *B. shortii*, *B. serotina* and *B. perstellata* constitute one lineage of eastern North American *Boechera* while *B. canadensis* constitutes a second one. This is also supported by chloroplast DNA data. However the relationships among those lineages and their relationship to western North American *Boechera* remain unclear. Therefore we suggest to retain *B. canadensis* and the group of *Boechera* species related to *B. laevigata* within *Boechera* instead of transferring them into different genera as has been suggested elsewhere. The Siberian *Boechera* species *B. falcata* and the Siberian genus *Borodinia* were shown to be placed within *Boechera*.

Phylogeographic analysis revealed a higher chloroplast diversity in southern regions for *B. canadensis* and a uniform haplotype distribution for *B. missouriensis* and *B. laevigata*. The route through which eastern North American *Boechera* arrived in their current habitats is unclear. Since the closest related haplotype AA/AB is only found in the southern Great Basin in the western United States either this haplotype had a wider distribution range into the east or eastern North American *Boechera* evolved closer to the distribution range of AA/AB, migrated into eastern North America and became extinct in the western parts of North America during the quaternary glaciation cycles.

LITERATURE CITED

Ahles, H.E. 1964. *Journal of the Elisha Mitchell Scientific Society* 80: 172.

Al-Shehbaz, I.A. (submitted) *Boechera* A. Löve and D. Löve. In: Flora North America.

Al-Shehbaz I.A., Beilstein, M.A. & Kellogg, E.A. 2006. Systematics and phylogeny of the *Brassicaceae* (*Cruciferae*): an overview. *Pl. Syst. Evol*. 259: 89--120

Al-Shehbaz, I.A. 2003. Transfer of most North American species of *Arabis* to *Boechera* (Brassicaceae). *Novon* 13: 381--391.

Al-Shehbaz, I.A. 2007. The North American genus *Sandbergia* (Boechereae, Brassicaceae). *Harvard Papers Bot.* 12: 425--427.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L., Warwick, S.I., Windham, M.D. & Al-Shehbaz, I.A. 2006. A Global nrDNA ITS Phylogeny of the Brassicaceae. *Mol. Biol. Evol.* 23: 2142--2160.

Bailey, C.D.M., Price, R. A. & Doyle, J. J. 2002. Systematics of the Halimolobine Brassicaceae: Evidence from Three Nuclear Loci and Morphology. *Syst. Bot*. 27: 318--332.

Beilstein, M.A., Al-Shebbaz, I.A. & Kellogg, E.A. 2006. Brassicaceae phylogeny and trichome evolution. *Am. J. Bot.* 93, 607--619.

Clement, M., D. Posada, & Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molec. Ecol*. 9: 1657--1659.

Dobeš, C., Mitchell-Olds, T. & Koch, M.A. 2004a. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A.* ×*divaricarpa*, and *A. holboellii* (Brassicaceae). *Molec. Ecol* 13: 349--370.

Dobeš, C., Mitchell-Olds, T. & Koch, M.A. 2004b. Intraspecific diversification in North American *Boechera* stricta (= *Arabis drummondii*), *Boechera* x divaricarpa, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. *Am. J. Bot.* 91: 2087—2101.

Dobeš, C., Koch, M.A., Sharbel, T.F. 2006. Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. *Annals Miss. Bot. Gardens* 93: 517--533.

Dobeš, C., Kiefer, C., Kiefer, M. & Koch, M.A. 2007. *Trn*F pseudogenes in North American genus *Boechera* (Brassicaceae): mechanistic aspects of evolution. *Pl. Biol*. 9: 1--14.

Doyle, J. J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem. Bull*. 19: 11--15.

Griffin S. R. & Barrett, S.C.H. 2004 Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: Inferences from Phylogeography. *Am. J. Bot.* 91: 465--473.

Godbout, J., Jamarillo-Correa, J.P., Beaulieu J. & Bousquet, J. 2005 A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molec. Ecol.* 14: 3497--3512.

Hewitt, G. 2000 The genetic legacy of the Quaternary ice ages. *Nature* 405: 907--913.

Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb III, T., Williams, J.W., & Hansen, B.C.S. 2000. Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quat. Science Rev*. 19: 489--508.

Kantama L., Sharbel T.F., Schranz M.E., Mitchell-Olds T., de Vries S. & de Jong, H. 2007 Diploid apomicts of the *Boechera holboellii* complex show large scale chromosome substitutions and different aberrant chromosomes. *Proc. Natl. Acad. Sci. USA* 104: 14026--14031.

Knight, C., Vogel, H., Kroymann, K., Shumate, A., Witsenboer, H., & Mitchell-Olds, T. 2006. Expression profiling and local adaptation of *Boechera holboellii* populations for water use efficiency across a naturally occurring water stress gradient. *Molec. Ecol*. 15: 1229--1237.

Koch, M.A., Dobeš, C., Kiefer, C., Schmickl, R. & Klimes, L. 2006. SuperNetwork identifies multiple events of plastid *trn*F (GAA) pseudogene evolution in the Brassicaceae. *Mol. Biol. Evol*. 24: 63--73.

Nicholas, K.B. & Nicholas, H. B. Jr. 1997. GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author.

Olson, D.M., Dinerstein, E., Wikramanayke, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. & Kassem, D.K.R. 2001. Terrestrial ecoregions of the world: New map of life on earth. *BioScience* 51: 933--938.

Magni, C.R., Ducousso, R., Petit, J.R. & Kremer, A. 2005. Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with Fagaceae. *Molec. Ecol*. 14: 513--524.

Rollins, R.C. 1993. The Cruciferae of Continental North America. Stanford University Press, Stanford.

Roy, B.A. 2001. Patterns of association between crucifers and their flower mimic pathogens: Host jumps are more common than coevolution or cospeciation.
Evolution 55: 41--53

Schmickl, R., Kiefer, C., Dobeš, C. & Koch, M.A. (in press) Evolution of *trn*F(GAA) pseudogenes in cruciferous plants. *Pl. Syst. Evol*.: doi:10.1007/s00606-008-0030-2.

Schranz, M.E., Dobeš, C., Koch, M.A. & Mitchell-Olds, T. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). *Am. J. Bot.* 92: 1797--1810.

Sharbel, T.F., Voigt, M.-L., Mitchell-Olds, T., Kantama, L. & de Jong, H. 2004. Are the aneuploid chromosomes in the apomictic *Boechera holboellii* complex B chromosomes? Camacho J.P.M. (ed): B Chromosomes in the Eukaryote Genome (special issue of Cytogenetic and Genome Research). *Cytogen. Genome Res.* 106(2-4): 173--183.

Shaw, J. & Small, R.L. 2005. Chloroplast DNA phylogeny and phylogeography of the North American Plums (Prunus subgenus Prunus section Prunucerasus, Rosaceae). *Am. J. Bot.* 92 (12): 2011--2030.

Siemens, D.H., Lischke, H., Maggiulli, N., Schürch, S. & Roy, B.A. 2003. Costs of resistance and tolerance under competition: the defense-stress benefit hypothesis. *Evol. Ecol*. 17: 247--263.

Smith, F.H. 1938. Some chromosome numbers in the Cruciferae. *Am. J. Bot.* 25: 220--221.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. 2006. Comparative phylogeography of unglaciated eastern North America. *Molec. Ecol*. 15: 4261--4293

Song, B.H., Clauss, M.J., Pepper, A. & Mitchell-Olds, T. 2006. Geographic patterns of microsatellite variation in *Boechera stricta*, a close relative of *Arabidopsis*. *Molec. Ecol*. 15: 357--369

Swofford, D.L. 2002. *PAUP*: Phylogenetic Anaylsis Using Parsimony (*and Other Methods)*, Version 4.0B10. Sinauer Associates, Sunderland, Massachusetts

Van de Peer, Y & De Wachter, R. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Applic. Biosci.* 10: 569-570.

Voigt, M.-L., Melzer, M., Rutten, T., Mitchell-Olds, T. & Sharbel, T.F. 2007. Gametogenesis in the apomictic *Boechera holboellii* complex: the male perspective. Pp. 235--258 in: Hörandl, E., Grossniklaus, U., can Dijk, P., Sharbel, T.F. (eds), *Apomixis: Evolution, Mechanisms and Perspectives*. International Association for Plant Taxonomy, Koeltz Scientific Books.

Warwick, S.I., Francis, A., Al-Shehbaz, I. 2006. Brassicaceae species checklist and database on CD-ROM. *Pl. Syst. Evol*. 259: 249--258.

Windham, M.D. & Al-Shehbaz, I.A. 2006. New and noteworthy species of *Boechera* I: sexual diploids. *Harv. Papers Bot.* 11: 61--88.

Windham, M.D. & Al-Shehbaz, I.A. 2007a. New and noteworthy species of *Boechera* (Brassicaceae) II: apomictic hybrids. *Harv. Papers Bot.* 11: 257--274.

Windham, M.D. & Al-Shehbaz, I.A. 2007b. New and noteworthy species of *Boechera* (Brassicaceae) III: additional sexual diploids and apomictic hybrids. *Harv. Papers Bot.* 12: 235--257.

Supplementary Table 1
Accession details of the plants included in the study and their haplotype information.

Supplementary Table 2
*Trn*LF alignment used for network analysis; sequence features are indicated above the alignment.

Supplementary Table 3
ITS-Alignment used for phylogenetic reconstruction; sequence features are indicated above the alignment.

Supplementary Table 4
*Trn*LF Alignment used for phylogenetic reconstruction; sequence features are given above the alignment.

3. Cytogenetic Analyses of apomictic *Boechera* (to be submitted to Heredity)

Manuscript to be submitted to Heredity.

Het-Hunting in *Boechera* apomicts – Searching for the Origin of a Heterochromatized and a Supernumerary Chromosome

Christiane Kiefer[1], Eric Schranz[2], Terezie Mandakova[3], Martin Lysak[3], Hans de Jong[4#]

[1] University of Heidelberg, Heidelberg Institute f Plant Sciences, Biodiversity and Systematics, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany
[2] University of Amsterdam, Instituut voor Biodiversiteit en Ecosysteem Dynamica; Kruislaan 318, 1098 SM Amsterdam, Netherlands
[3] Masaryk University, Department of Experimental Biology, Kamenice 5, 62500 Brno, Czek Republic
[4] Wageningen University, Laboratory of Genetics, Arboretumlaan 4, 6703 BD Wageningen, the Netherlands

Corresponding author:
Hans de Jong; Wageningen University, Laboratory of Genetics, Arboretumlaan 4, 6703 BD Wageningen, the Netherlands. Tel. +31 629023329; e-mail. hans.dejong@wur.nl

Running Title: Heterochromatic Chromosome in *Boechera* Apomicts

Keywords: *Boechera*, apomixis, heterochromatic, cytogenetics

Wordcount:

Abstract

Usually apomixis, a mode of asexual reproduction, is associated with higher ploidy levels among plants. Apomixis is a frequent mode of reproduction among North American *Boechera*, however, in *Boechera* apomixis on the diploid and triploid level is common. Past cytogenetic studies revealed that some apomictic *Boechera* accessions are aneuploid and have a small extra chromosome which was assumed to be a B chromosome. Furthermore it was shown that all diploid and aneuploid apomictic accessions have a heterochromatized (Het) chromosome which pairs with another chromosome in meiosis. In a BAC-FISH appproach we reveal that the Het chromosome is a homologue of *Boechera stricta* linkage group 1 which is missing a fragment (genome block D) in 15 chromosome apomicts. The missing fragment was apparently transfered to a triplicate centromere or has fomed a neo-centromere by itself. The fragment corresponds to the previously described Del chromosome. Results from the present BAC-FISH analysis are embedded into results from past cytogenetic analyses. We can show that *Boechera* does not have a B chromosome.

Introduction

Meiotic cell division in sexual reproducing organisms requires an even chromosome number and pairs of homologous chromosomes for a functional first division. Therefore large scale chromosome mutations, for example translocations or inversions, will be “punished” by rendering the individual who carrying such a mutation infertile or of reduced fertility. However, reproduction may also occur asexually which means that meiosis is not longer required and embryos may develop spontaneously from a somatic cell or an embryo-sac mother cell (Bicknell and Koltunow, 2004). This way of reproduction is refered to as apomixis and is rarely found in the animal kingdom but evolved several times independently throughout the plant kingdom (Bicknell and Koltunow, 2004). The lack of meiotic cell division means that uneven chromosome numbers and large scale chromosome mutations can be tolerated.

In the plant kingdom large apomictic complexes are found within the Asteraceae (*Taraxacum* (e.g. van Dijk et al. 1999); *Hieracium* e.g. Eckardt, 2003), the Ranunculaceae (*Ranunculus cassubicus* complex, Paun et al. 2006) and the Rosaceae (*Potentilla* e.g. Nylehn et al. 2003; *Rubus* e.g. Nybom, 1988). However, although absent in almost all members of the family apomictic reproduction is very common in one genus of the Brassicaceae, the genus *Boechera* from North America (Sharbel et al. 2001, Dobes et al. 2006, Dobes et al. 2007). According to the current taxonomic circumscription it comprises 110 species of which 38 are apomictic. Apomicts in *Boechera* are assumed to be of hybrid origin with various combinations of parents (Sharbel and Mitchell-Olds, 2001, Windham and Al-Shehbaz 2006, 2007a, 2007b). They are pseudogamous and diplosporous which means that fertilization of the endosperm is still required and the embryo is formed from an unreduced embryo-sac morther cell (Naumova et al. 2001). Past studies revealed that *Boechera* apomicts are facultative apomicts which means that to a variable extent they can still take part in sexual reproduction and cross back into sexual populations (Böcher, 1951, Dobes et al. 2006, Schranz et al. 2005). Up to date it is unclear what effect this behaviour has on the sexual reproducing populations and if the potential to become apomicitc can be transferred that way.

Apomicts in *Boechera* are mostly triploid but diploid or aneuploid apomicts also exist (Sharbel and Mitchell-Olds, 2001, Dobes et al. 2006). The latter were subject to cytogenetic investigations in the past which revealed the presence of translocations (one individual, BH115, with a translocation complex visible during metaphase I in meiosis) as well as one supernumerary small chromosome referred to as Del (deletion chromosome) in 15 chromosome apomicts and a heteropycnotic chromosome referred to as Het (heterochromatized chromosome) found in all investigated apomicts (Kantama et al. 2007).

Another characteristic feature was that apomicts had different numbers of parental chromosomes indicating their ability to occasionally reproduce sexually and therebye recombining their chromosome sets (Kantama et al. 2007).
The Del chromosome was shown to align with another chromosome pair during meiosis forming a heteromorphic trivalent which suggested a triplication of at least a portion of the Del chromosome. Its behaviour fitted that of a translocation trisomic.
The Het chromosome formed a heteromorphic bivalent during meiosis suggesting that at least parts of it are homologous to another chromosome. The Het chromosome was heteropycnotic in its behaviour meaning that it showed the opposite behaviour in contraction and extension compared to the remaining chromosomes in the chromosome set (Kantama et al. 2007). With its heteromorphic appearance and heteropycnotic behaviour the Het chromosome seemed to be comparable to the human Y chromosome (personal communication Hans de Jong).
In the past it was described that the supernumerary chromosome in *Boechera* is a B chromosome (Sharbel et al. 2004, 2005).
Since chromosomes in *Boechera* are comparably small (3-6 µm on average) and do not display obvious features which can be identified by eye differentiation of chromosomes in phase contrast or after DAPI staining in the fluorescence microscope is impossible. Therefore it is not known to what other chromosome or to what part of another chromosome Het and Del are homologous. Since the Het chromosome has been only identified in apomictic accessions one feels tempted to hypothesize that the Het chromosome is involved in the expression of apomixis. However, proof for this is still lacking, but the identification of the origin of the Het chromosome would be a first step into the direction of revealing its role in apomixis if any. Since the Del chromosome is also exclusively found in apomicts the identification of its origin would be helpful for the characterization of aneuploid 15 chromosome apomicts in Boechera.
For the identification of the origin of Het and Del as well as the detection of further transloctions we chose a BAC-FISH approach. The genome of the Brassicaceae is arranged in genome blocks. This genome blocks are syntenic arrays of genes which are shuffeled as blocks that occur throughout the Brassicaceae family (Schranz et al. 2006). Genetic studies of *Boechera stricta* indicated the most likely position of genome blocks in the Boechera genome (Schranz et al. 2007). With repeat free Arabidopsis thaliana BACs repesenting the genomic blocks we aim to identify (1) the origin of the Het chromosome, (2) the origin of the Del chromosome and (3) the presence of translocations within genomes of different diploid and aneuploid Boechera apomicts.
.

Materials and Methods

Plant Material

Apomictic and sexual plants were grown in the greenhouse under the natural light rhythms from April until August. Plants were put into new soil every two months and fertilized every second week to enhance root growth. The day before root tip collection the plants were lifted out of their pots to enhance growth of fresh root tips. Root tips were harvested between 8 and 9 a.m. and stored directly in 2 mM 8-hydroxychinolin for 3h at 15 °C. Afterwards they were fixed in Carnoy fixative (ethanol-acetic aced 3:1) for at least one hour or overnight. Flower buds were collected between 10 and 11 a.m. and directly fixed in Carnoy fixative.
Details on the origin of the plants and results from previous studies can be found in Table 1.

Table 1. Accession Details of plant material used in this study

Accession	geographical origin	chromosome number	Het present	Del present	*B. holboellii* chromosomes	*B. stricta* chromosomes
BH74.40	Ranch Creek, Granite Co., MT, USA	15	yes	yes	4	11
BH115	Birch Creek, Ravalli Co., MT, USA	15	yes	yes	5	10
ES9	Vipond Park, Beaverhead Co., MT, USA	14	yes		7	7

Metaphase and Meiotic Chromosome Preparations

Root tips and flower buds were washed in milli-Q water for 5 min at room temperature three times and then left in 10 mM citrate buffer pH=4.5 for 5 min to equilibrate. Root tips were digested in 0.02% enzyme mix (0.3% Cellulase RS, 0.3% Pectolyase, 0.3% Cytohelicase) in the same citrate buffer for 1-2h at 37°C. Flower Buds were digested in 0.1% of the enzyme mixture also in citrate buffer as described above for 1-2h at 37°C. Afterwards the material was stored on ice. One to five root tips or one flower bud were placed on a slide and excess liquid was removed. Then the material was tapped with a dissection needle resulting in a homogeneous cell suspension. A drop of 50% acetic acid was added and the material was stirred carefully until no cytoplasm was visible any more. Then the slide was incubated on a hot plate at 45 °C for 1-2 min with occasional careful stirring of the acetic acid droplet. For fixing Carnoy fixative was dropped around and then on top of the acetic acid droplet and left for a few seconds before the slide was left to dry on the hot plate. Quality of the slides was assessed under the phase contrast microscope at 40x magnification.

Table 2 *Arabidopsis thaliana* BACs used in this study and their position in the *Boechera stricta* genome derived from to the genetic map by Schranz et al. 2007.

BAC name	genome block	chromosome in *Boechera stricta*	BAC name	genome block	chromosome in *Boechera stricta*
T25K16	A	LG 1	**T10F5**	H	LG 5
F21M11	A	LG 1	**T6B13**	H	LG 5
F14J9	A	LG 1	**F6P23**	H	LG 5
F2J6	C	LG 1	**F19F24**	H	LG 5
F11A17	C	LG 1	**F23N11**	H	LG 5
T18C15	C	LG 1	**F2I9**	K	LG 5
F8L10	C	LG 1	**F5O4**	K	LG 5
F13N6	D	LG 1	**F3C11**	K	LG 5
T18I24	D	LG 1	**MJL14**	L	LG 5
T30E166	D	LG 1	**MMJ24**	L	LG 5
T25B24	D	LG 1	**MFJ20**	L	LG 5
T12P18	D	LG 1	**MTO24**	L	LG 5
F5M15	B	LG 2	**T4A2**	L	LG 5
T23E23	B	LG 2	**T10D17**	M	LG 5
F28L5	B	LG 2	**F14D17**	M	LG 5
T19E23	B	LG 2	**T6H20**	M	LG 5
F12K21	B	LG 2	**T21J18**	M	LG 5
T23K8	E	LG 2	**F3A4**	M	LG 5
T6L1	E	LG 2	**F24M12**	N	LG 5
F14O23	E	LG 2	**F28P10**	N	LG 5
F1B16	E	LG 2	**F24I3**	N	LG 5
F23A5	E	LG 2	**T2O9**	N	LG 5
T4P13	F	LG 3	**F16M2**	N	LG 5
F24K9	F	LG 3	**F5I10**	O	LG 6
MVC8	F	LG 3	**F3D13**	O	LG 6
MOE17	F	LG 3	**TsH3**	O	LG 6
MWL2	F	LG 3	**T27D20**	O	LG 6
F16J10	G	LG 3	**T1J1**	O	LG 6
T6P5	G	LG 3	**T3H13**	P	LG 6
T25N22	G	LG 3	**T5L19**	P	LG 6
K21P3	W	LG 3	**F8L21**	P	LG 6
MIO24	W	LG 3	**F7J8**	R+Q	LG 6
K19P17	W	LG 3	**MHF15**	R+Q	LG 6
MHM17	W	LG 3	**T30N20**	R+Q	LG 6
K9B18	W	LG 3	**T19L5**	R+Q	LG 6
MSL3	X	LG 3	**T21H19**	R+Q	LG 6
K9H21	X	LG 3	**F17K4**	R+Q	LG 6
MVP7	X	LG 3	**F22D1**	R+Q	LG 6
K1F13	X	LG 3	**T32G24**	R+Q	LG 6
K9I9	X	LG 3	**T4C12**	R+Q	LG 6
F7O24	I	LG 4	**T19G15**	R+Q	LG 6
T20D16	I	LG 4	**F5H8**	S	LG 7
F27A10	I	LG 4	**MXA21**	S	LG 7
T19L18	I	LG 4	**MUL8**	S	LG 7
T16B12	J	LG 4	**MNF13**	S	LG 7
F11F19	J	LG 4	**MPK23**	S	LG 7
T2P4	J	LG 4	**F25E4**	T	LG 7
F23E6	J	LG 4	**T20K18**	T	LG 7
T8I3	J	LG 4	**F18A5**	T	LG 7
MBD2	V	LG 5	**T6K21**	U	LG 7
MQO24	V	LG 5	**F1N20**	U	LG 7
K15C23	V	LG 5	**F20O9**	U	LG 7
F10E10	V	LG 5	**F8D20**	U	LG 7
K23F3	V	LG 5	**T5J17**	U	LG 7

BAC isolation and Labelling

BACs were grown over night at 37 °C from glycerol stocks in 4 mL LB medium with either chloramphenicol (12.5 µg/ml LB) or kanamycin (10 µg/mL LB). Isolation of BACs followed the protocol supplied with the High Pure plasmid isolation kit (Roche diagnostics).

BACs were labelled block wise either indirectly with Biotin (later to be detected with avidin-TexasRed or Cy5) or DIG (later to be detected with anti-DIG-FITC) or directly with Cy3 using nick translation mix supplied by Roche diagnostics (catalogue numbers 11 745 824 910, 11 745 816 910, 11745808910).

Fluorescence In Situ Hybridisation

The FISH procedure followed essentially the protocol outlined in Lysak et al. 2006. Changes included the following: the washing step after the hybridisation was done with 50% formamide in SSC for 5 min in each step at 42 °C and the slide was incubated with 100 µl TNB for 30 min at 37 °C prior to the incubation with the antibodies to achieve a better blocking.. For identity of BACs and the block to which they were assigned refer to table 2.

Results

The Het and Del Chromosome in the 15 chromosome apomict BH74.40

In BH74.40 two copies of chromosomes 2, 3, 4, 5, 6, and 7 were detected and the pattern of blocks was congruent with that described by Schranz et al. 2007 (figure 1 b-f showing LG2, LG3, LG4, LG5 in two parts). One copy of chromosome 1 was identified which showed the three blocks A, C and D (figure 1a). The Het chromosome was easily distinguishable from the other chromosomes by its much brighter DAPI staining after the FISH procedure and showed signals of block A and C from chromosome 1. However, the second signal for block D was found apart from blocks A and C on a very small chromosome (figure 1a).

Three signals for block U from chromosome 7 were detected. The signal for block U did not occur together with the signal for block D on the small chromosome.

A cartoon of the karyotype of BH74.40 representing the major findings from the FISH analysis is given in figure 2a.

The Het and Del chromosome in the 15 chromosome apomict BH115

In BH115 the Het chromosome was homologous to blocks A and C from chromosome 1. Block D was found apart from blocks A and C on the Het chromosome on a small chromosome (figure 1 g). Figure 2b displays a cartoon of chromosome pairs 1 to 4.

The Het chromosome in the14 chromosome apomict ES9

The Het chromosome displayed the signals for block A, C and D from chromosome 1. There was no translocation of block D to a different chromosome (figure 1 h). Figure 2c displays a cartoon of chromosome pairs 1, 2, 3, 4, 6, and 7.

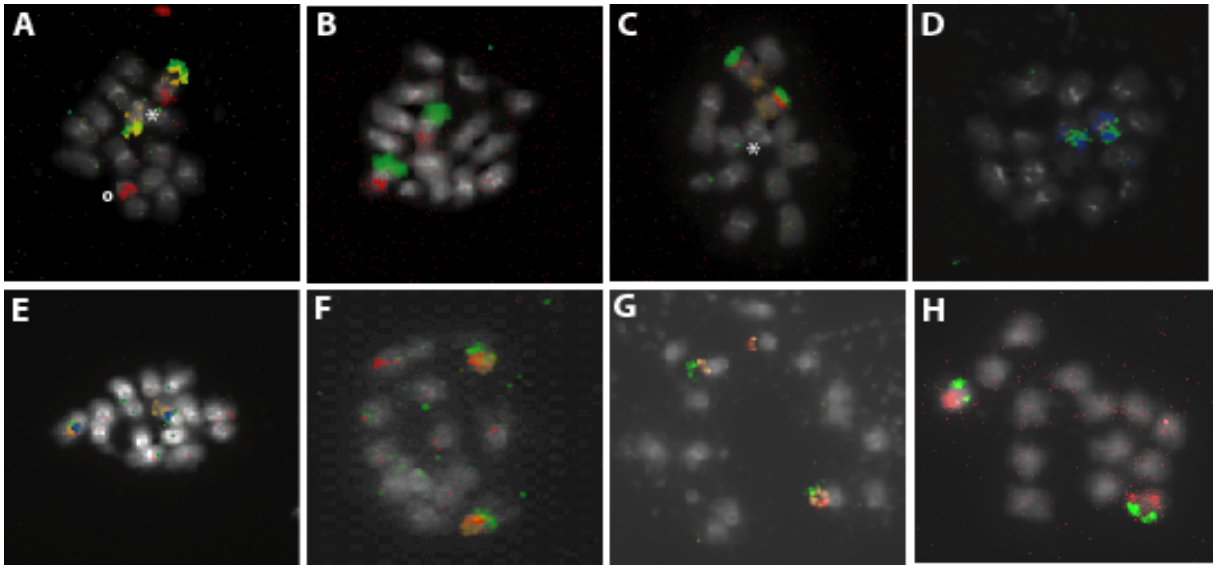

Figure 1 Examples fluorescent microscopy images of BAC-FISH analyses A BH74.40 LG1, The Het chromosome has only signals for blocks A (green) and C (orange), block D is located apart from it on a small extra chromosome, the second LG1 shows signals for blocks A, C and D; B BH74.40 LG2; C BH74.40 LG3; D BH74.40 LG4, E and F BH74.40 LG5 upper and lower section with overlap; G LG1 BH115; the same pattern as in LG1 of BH74.40 is seen; H LG1 of ES9; blocks A, C and D are located on the Het chromosome.

Discussion

Supernumerary chromosomes in *Boechera* were already described as early as 1951 in the elaborate studies of Tyge Böcher (Böcher, 1951), the cytologist whose name was later given to the genus into which the Greenlandic *Arabis* species *Arabis holboellii* was transferred (Löve and Löve, 1976). Later studies investigated the nature of supernumerary chromosomes in 2n + 1 and 3n apomicts closer (Sharbel et al. 2004, Sharbel et al. 2005). But the findings from those studies were contradictory when the nature of the additional chromosome, an assumed B chromosome, was described. In the first publication the additional chromosome was thought to be heterochromatic while later it was identified as a chromosome of small size. It was thought that the supernumerary chromosome does not pair with a homologue during meiosis. A cytogenetic study of *Boechera* apomicts finally described the heterochromatic chromosome as the Het chromosome and the small chromosome as the Del (from deletion) chromosome (Kantama et al. 2007). The Het chromosome was present in all 2n and 2n + 1 apomicts while the Del chromosome was only present in the 2n + 1 individuals. Unlike the findings from previous studies both Het

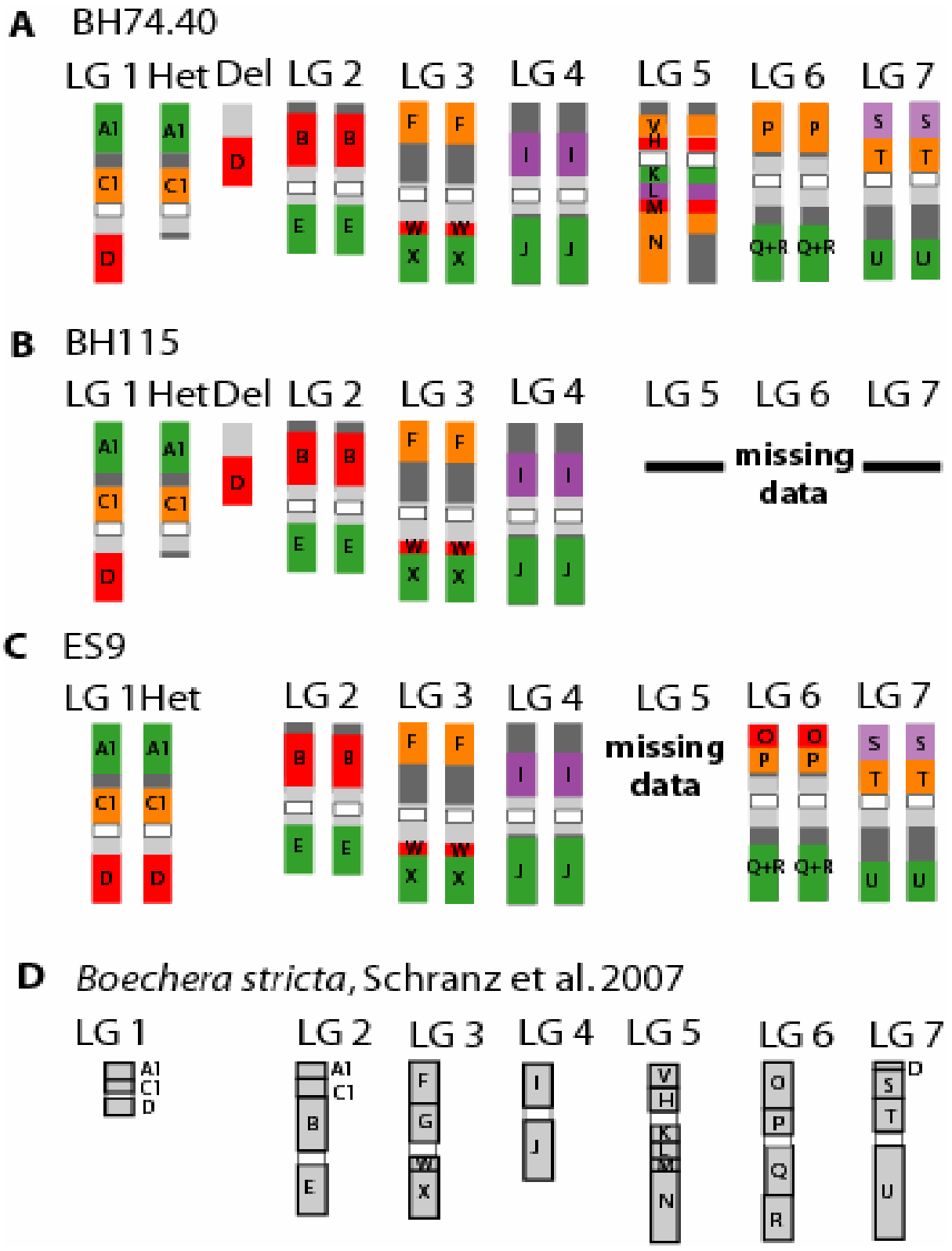

Figure 2 Drawings of the identified orders of blocks on the linkage groups of *Boechera* apomicts in comparison to the genetic map (Schranz et al. 2007). A BH74.40, B BH115, C ES9, D order and distribution of blocks according to the genetic map

and Del were shown to pair with other chromosomes during meiosis (however, not shown by painting experiments). The Het formed a heteromorphic bivalent while Del

was associated with two other chromosomes. Genome Painting experiments suggested that the Het and the Del chromosome originated from *Boechera stricta*.

A *Boechera stricta* origin of the supernumerary chromosome was also assumed from the B chromosome studies (Sharbel et al. 2004, Sharbel et al. 2005). In those studies the evidence was taken from the presence of a third group of alleles for the markers MVI11 and T1B9 which clustered with alleles typical for *Boechera stricta*. Both markers are localized on *Arabidopsis chromosome* 3. Comparing their position with the genetic map of *Boechera stricta* and the Brassicaceae genome blocks (Schranz et al. 2007) both markers should be located in block F on *Boechera stricta* linkage group 3 (Brassicaceae genome blocks as described in Schranz et al. 2006).
Our central aim in this study was to elucidate the origin of the Het and the Del chromosome. Based on the genetic map for *Boechera stricta* (Schranz et al. 2007) we applied a BAC-FISH approach.
The Het chromosome was easily identified by its bright DAPI stain. In the 2n +1 apomicts BH115 and BH74.40 the signal for genome block A and C were found on the Het chromosome. Block D was found to be apart on a suspiciously small chromosome which we identified as the Del chromosome. In the investigated 2n apomict ES9 the signals for block A, C and D were present on the Het chromosome. According to the genetic map of *Boechera stricta* (Schranz et al. 2007) the blocks A, C and D are located on *Boechera stricta* linkage group 1. In the genetic map parts of block A and C were also located on linkage group 2. However, we used only BACs which would bind to the portions of blocks A and C on linkage group 1.
From our findings we can conclude that the Het chromosome is homologous to *Boechera stricta* linkage group 1 in both 2n and 2n + 1 apomicts. However, in 2n + 1 apomicts the Het chromosome consists only of two out of three blocks (A and C). The third block, block D, was either translocated to another chromosome which then was degraded to its small size or it broke off the heterochromatic chromosome, formed a neo-centromere and became a chromosome by itself. This small chromosome is the Del chromosome.
Since the Del chromosome was shown to pair with two other chromosomes it is more likely that it is a fusion to a triplicated centromere than a small chromosome by itself with a neo-centromere although this possibility cannot be ruled out. Neo-centromeres formation exists in the plant kingdom (Dawe, 2005) and spontaneous breakage of heterochromatic chromosomes or heterochromatic regions is common(Sacristan, 1971, McCoy et al. 1982, Fluminhan et al. 1996).
Our knowledge on the nature of the Het and Del chromosome together with the findings of Kantama et al. 2007 rules out that those two heteromorphic chromosomes in *Boechera* are B chromosomes. First of all the Het chromosome is simply a homologue of LG1/a part of LG1, the only difference being that it is heterochromatized. The Del chromosome pairs with another chromosome pair at

meiosis indicating a potential triplication. Pairing with other chromosome is against the definition of a B chromosome. Furthermore, the statement that the "B chromosome" seems to be dispensable, which is also a criterion in the definition of B chromosomes, is not necessarily true. The Del chromosome which was interpreted to be the B chromosome in Sharbel et al. 2005 simply does not exist in 14 chromosome apomicts. The block which makes up the Del chromosomes is in place in *Boechera stricta* linkage group 1 in 2n apomicts. . In other words, no large scale deletions exist in these accessions. From these finding we henceconclude that the Het and Del chromosomes should no longer be referred to as B chromosomes.

In our study we could not find any indication on what caused the heterochromatinization of the Het chromosome. In *Drosophila miranda* the formation of neo-Y chromosomes was investigated (Steinemann and Steinemann, 2008). The authors came up with a model in which the degeneration of the neo-Y chromosome resulted from an accumulation of point mutations and massive spread of transposons, especially retrotransposons. In the *Drosophila* model transposons could accumulate because recombination of the X and Y chromosome was suppressed. The enrichment of transposable elements finally resulted into heterochromatinization. This *Drosophila* model can give us an idea of what may have happened in *Boechera*. Recombination in the pair of linkage group 1 of *Boechera stricta* is suppressed for an unknown reason (Schranz et al. 2007). This is the same situation as in *Drosophila*. Past genome painting experiments in *Boechera* suggested the presence of unique repeats in the Het chromosome (Hans de Jong, personal communication). Following the *Drosophila* model they could be the reason for the heterochromatinization of one of the LG1 chromosomes in *Boechera*.

Genome subtraction would give us the possibility to find apomict specific sequences. Further FISH experiments would enable us to see if they are really located on the Het chromosome and sequence identity may reveal something about their nature, for example if they originate from transposable elements.

Apomixis has often been described as a dosage effect (e.g. Grimanelli et al. 1997) rather than resulting from a particular mutation in a gene. If heterochromatinization silences genes in one of the copies of LG1 it will also produce a dosage effect. In this scenario the Het chromosome would have an indirect effect on the expression of apomixis by halving the amount of a gene product required for sexual reproduction.

References

Bicknell, R. A., Koltunow, A. M. 2004. Understanding Apomixis: Recent Advances and Remaining Conundrums. The Plant Cell 16: S228-S245.

Böcher, T. W. 1951. Cytological and embryological studies in the amphi-apomictic *Arabis holboellii* complex. Kong Danske Vidensk Selsk, 6: 1–59.

Dawe, R. K. 2005. Centromere renewal and replacement in the plant kingdom. PNAS 102: 11573-11574.

Dobeš, C., Koch, M.A., Sharbel, T.F. (2006) Embryology, karyology, and modes of reproduction in North American genus *Boechera* (Brassicaceae): a compilation of seven decades of research. Annals Miss. Bot. Gard. 93: 517-533.

Dobeš, C., Sharbel, T.F, Koch, M.A. (2007) Towards understanding the dynamics of hybridization and apomixis in the evolution of genus *Boechera* (Brassicaceae). Systematics and Biodiversity 5 (3): 321-332.

Eckardt, N. A. (2003) Patterns of Gene Expression in Apomixis. The Plant Cell, Vol. 15, 1499-1501.

Fluminhan, A. Jr., de Aguiar-Perecin, M. L. R., Dos Santos, J. A. 1996. Evidence for Heterochromatin Involvement in Chromosome Breakage in Maize Callus Culture. Ann Bot 78: 73-81.

Grimanelli, D., Hernández, M., Perotti, E., Savidan, Y. 1997. Dosage effects in the endosperm of diplosporous apomictic Tripsacum (Poaceae). Sexual Plant Reproduction 10: 279-282.

Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., de Vries, S., de Jong, H., 2007. Diploid apomicts of the *Boechera holboellii* complex show large scale chromosome substitutions and different aberrant chromosomes. Proceedings of the National Academy of Sciences 104, 14026-14031.
Löve, A. and D. Löve. 1976. Botaniska Notiser 128: 513.

Lysak, M., Frasz, P., Schubert, I. 2006 in Methods in Molecular Biology, vol. 323, p. 173-186: Arabidopsis Protocols, Second Edition. Edited by: J. Salinas and J. J. Sanchez-Serrano, Humana Press Inc., Totowa, NJ

McCoy TJ, Phillips RL, Rines HW. 1982. Cytogenetic analysis of plants regenerated from oat (*Avena sativa*) tissue cultures ; high frequency of partial chromosome loss. Canadian Journal of Genetics and Cytology 24: 37-50.

Naumova, T.N., van der Laak, J., Osadtchiy, J., Matzk, F., Kravtchenko, A., Bergervoet, J., Ramulu, K. S., Boutilier, K. 2001 Reproductive development in

apomictic populations of *Arabis holboellii* (Brassicaceae). Sexual Plant Reproduction 4: 195-200

Nybom, H. Apomixis versus sexuality in blackberries (*Rubus* subgen.Rubus, Rosaceae) 1988. Plant Systematics and Evolution 3: 207-218.

Nylehn, J., Hamre, E., Nordal, I. 2003. Facultative apomixis and hybridization in arctic *Potentilla* section Niveae (Rosaceae) from Svalbard. Botanical Journal of the Linnean Society 142: 373-381.

Paun, O., Stuessy, T. F., Hörandl, E. (2006) The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex. New Phytologist 171: 223-236.

Sacristan, M.D. 1971. Karyotypic changes in callus cultures from haploid and diploid plants of *Crepis capillaris* (L.) Wallr. Chromosoma 33: 273-283.

Schranz, M. E., Lysak, M. A., Mitchell-Olds, T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. TRENDS in Plant Science 11: 535-542.

Schranz, M. E., Windsor, A. J., Song, B.-H., Lawton-Rauh, A., and T. Mitchell-Olds. 2007. Comparative Genetic Mapping in *Boechera stricta*, a Close Relative of Arabidopsis. Plant Physiology 144: 286–298.

Sharbel, T. F., Mitchell-Olds, T. (2001) Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). Heredity 87: 59-68.

Sharbel, T.F., Voigt, M.-L., Mitchell-Olds, T., Kantama, L., de Jong, H. (2004) Is the aneuploid chromosome in apomictic *Boechera holboellii* a genuine B chromosome? Cytogenetic Genome Research 106: 173-183.

Sharbel, T.F., Mitchell-Olds, T., Dobeš, C., Kantama, L., de Jong, H. (2005) Biogeographic distribution of polyploidy and B chromosomes in the apomictic *Boechera holboellii* complex. Cytogenetic Genome Research 109: 283-292.

van Dijk, Peter J, Tas, Inge C Q, Falque, M., Bakx-Schotman, T. 1999. Crosses between sexual and apomictic dandelions (Taraxacum). II. The breakdown of apomixis. Heredity 83(6):715-721.

Windham, M.D., Al-Shehbaz, I.A., 2006. New and noteworthy species of *Boechera* (Brassicaceae) I: sexual diploids. Harvard Papers in Botany 11, 61-68.

Windham, M.D., Al-Shehbaz, I.A., 2007a. New and Noteworthy Species of *Boechera* (Brassicaceae) II: Apomictic Hybrids. Harvard Papers in Botany 11, 257-274.

Windham, M.D., Al-Shehbaz, I.A., 2007b. New and Noteworthy Species of *Boechera* (Brassicaceae) III: Additional Diploids and Apomictic Hybrids. Harvard Papers in Botany 12, 251-274.

4. Overview of Scientific Contributions

Koch, M., Kiefer, C. 2006. Molecules and migration: biogeographical studies in cruciferous plants. Plant Systematics and Evolution 259: 121-142.

Koch, M., Kiefer, C., Vogel, J., Ehrich, D., Brochmann, C., Mummenhoff, K. 2006. Three times out of Asia Minor - the phylogeography of *Arabis* alpina L. (Brassicaceae). Molecular Ecology; 15: 825-839.

Koch M., Dobeš, C., Kiefer, C., Schmickl, R., Klimes, L., Lysak, M. A. 2007. Plastidic trnF (GAA) pseudogene evolution indicates an ancient split in the phylogeny of the Brassicaceae. Molecular Biology and Evolution 24: 63-73.

Dobeš Ch., Koch M., Kiefer, C. 2007. Apomixis and radiation at low ploidy levels exemplified in the evolutionary model genus *Boechera* (Brassicaceae). Regnum Vegetabile 147: 391-412.

Dobeš C., Kiefer, C., Kiefer M., Koch, M. 2007. SuperNetwork identifies multiple events of plastid trnF (GAA) pseudogene evolution in the Brassicaceae. Plant Biology 9: 502-515.

Calonje M., Martin-Bavo S., Dobeš C., Gong W., Jordon-Thaden I., Kiefer C., Kiefer M., Paule J., Schmickl R., Koch M. A. (2008) Non-coding nuclear DNA marker in phylogenetic reconstructions. Pl. Syst. Evol. Pl. Syst. Evol. 10.1007/s00606-008-0031-1

Schmickl R., Kiefer C., Dobeš C., Koch M. A. (2008) Evolution of trnF(GAA) pseudogenes in cruciferous plants. Pl. Syst. Evol. DOI 10.1007/s00606-008-0030-2.

6. Contents supplementary material and raw data on the DVD included in the thesis

Accessionlist

- Accessionlist.mdb
- Accessionlist.xls

Pictures Herbarium Specimens

- DAO and GH
- MO

Raw Data *(ordered in sub folders according to accession number)*

- chloroplast markers
- nrDNA
- nuclear intron sequences

Sampling_Map_all_Accession.jpg

Supplementary Data Chapter 1
includes all supplementary data mentioned in the text in chapter 1

Supplementary Data Chapter 2
includes all supplementary data mentioned in the text in chapter 2

Supplementary Data Chapter 3
includes all supplementary data mentioned in the text in chapter 3

Thesis_Christiane_Kiefer_2008.pdf

5. Acknowledgements

A PhD thesis is a big challenge and as in almost every situation in life success depends on teamwork rather than on being a lone wolf. A number of people have contributed to the success of my research and also to my personal development over the past three years. These great personalities shall be mentioned in the following.

First of all I would like to thank Prof. Marcus Koch (HIP, University of Heidelberg), who gave me the opportunity to work in his lab prior to my time as a PhD student and then in this challenging project for my PhD thesis. Prof. Marcus Koch has not merely been an excellent PhD supervisor but also a friend in tough situations.

I would like to thank Prof. Thomas Rausch (HIP, University of Heidelberg) for being the second referee. Prof. Thomas Rausch roused my interest in botany and played a big role in my molecular biological education together with his group.

I would like to express my gratitude towards Prof. Hans de Jong (WUR/NL) and Dr. Eric Schranz (University of Amsterdam/NL) who made it possible for me to spend five month in the Netherlands in the framework of a DAAD scholarship.

I would also like to express my gratitude and admiration towards Dr. Ihsan Al-Shebaz who did this amazing work on the taxonomy of *Boechera* and innumerable other Brassicaceae.

Furthermore I would like to say thanx to everyone in the labs of Marcus Koch and Hans de Jong and Claudia Erbar– you have been great colleagues! Without you it would have been half the fun...

A special thanx goes to Wilma Lasthuis from Wageningen (Dank je wel... Jij bent een echt goede vriendin!), Dora Szinay from the Lab of Hans de Jong (take care of that material, my friend!), Roswitha Schmickl from the Koch Lab (we will always be the unforgettable Duo Infernale) and Andrea Schott from the Strahl Lab and Astrid Ledebuhr now a gardener in England (wonderful girls' nights!!!!)

Another very special thanx goes to Dr. Markus Kiefer (who is completely unrelated to me) and Dr. Steffen Greiner for many great lunch times, coffee breaks and computer problem solutions (this goes to Markus) and all the jokes. I can't imagine to invent the concept of a portable black hole with anybody else....
And Dr. Peter Sack for being a very good friend... thanx for taking care of me.

Last but not least I would like to mention those persons, who are dearest to me in my life: my family. These are my parents Erich and Karin Kiefer who made me the person I am and my brothers Lars-Erich, Mirko and Boris and my aunt and her family. Tough times tied us together – we are an amazing team!
Together with my family I want to thank Rafael Martinez, who came into my life with a joke. Thank you for being with me now and ever....