Spring 2018

# Improving the Phylogenetic Understanding of the Genus Juniperus

Therese Balkenbush
*Central Washington University*, balkenbust@cwu.edu

# IMPROVING THE PHYLOGENETIC UNDERSTANDING OF THE GENUS *JUNIPERUS*

---

A Thesis

Presented to

The Graduate Faculty

Central Washington University

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Biology

---

by

Therese Balkenbush

May 2018

CENTRAL WASHINGTON UNIVERSITY

Graduate Studies

We hereby approve the thesis of

Therese Balkenbush

Candidate for the degree of Master of Science

APPROVED FOR THE GRADUATE FACULTY

_____ _____
Dr. Linda Raubeson, Committee Chair

_____ _____
Dr. Jim Johnson

_____ _____
Dr. David Darda

_____ _____
Dean of Graduate Studies

ABSTRACT

IMPROVING THE PHYLOGENETIC UNDERSTANDING OF THE GENUS *JUNIPERUS*

by

Therese Balkenbush

May 2018

*Juniperus* is a species-rich and geographically widespread genus of coniferous trees and shrubs. The genus is relatively recently diverged, and has experienced periods of rapid diversification. Recent phylogenetic investigations by others have compared DNA from selected regions of the chloroplast, but the resulting topologies conflict, and some relationships remain unresolved. Their relatively small data sets failed to capture sufficient variation to resolve events of rapid diversification in these closely related taxa. This study provides increased resolution and support by generating a plastome-scale phylogeny for 28 *Juniperus* species, revealing previously unresolved relationships at both deep and shallow nodes. One-third of the recognized species of *Juniperus* are included, representing each of the major clades within the genus. This study contributes eight complete and 17 nearly complete chloroplast genome sequences to an ever-growing number of sequenced organellar genomes. This phylogeny provides a foundation from which an improved biogeographic history and molecular dating analysis can be performed.

# ACKNOWLEDGEMENTS

I'd like to thank my graduate advisor, Dr. Linda Raubeson of Central Washington Univeristy, who provided steady and patient guidance during my graduate studies. She took a chance, and gave me the opportunity to pursue a subject matter that had fascinated me, but of which I had minimal prior knowledge. I'd also like to thank my graduate committee members Dr. Jim Johnson and Dr. Dave Darda, who agreed to participate on my committee and provide guidance and feedback at each phase of my graduate work. Rhiannon Peery (CWU) and Julie Morris (UT Brownsville) prepared the samples to generate the raw data for the 17 partial genomes, funded by NSF award DEB 0629890 to Linda Raubeson and NSF award DEB 0629402 to Andrea Schwarzbach. Dr. Robert Adams contributed juniper specific expertise during correspondence. Funding for the current work was awarded by the School of Graduate Studies of Central Washington University in the form of the Graduate Student Summer Fellowship and the Research/Activity Award. I'd like to acknowledge the Denver Botanic Gardens, the Montgomery Botanical Center, and the Arnold Arboretum for their contribution of plant tissue samples from taxa included in this study. Lastly, I'd like to thank my parents and husband for their patience, love, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

*Juniperus* is the largest of 32 genera in the conifer family Cupressaceae, with 75 recognized species of trees and shrubs (Adams and Schwarzbach 2013b). In addition to high species diversity, the genus inhabits a broad geographic and topographic range. With one exception, *Juniperus* species are found exclusively in the Northern Hemisphere, at elevations ranging from sea level to above timberline (Adams 2014). Additionally, they have successfully colonized environments ranging from deserts to bogs and oceanic islands, and a variety substrates and soil types. Despite the variety of landscapes inhabited, many juniper species are primary components of tree/shrub vegetation communities in arid and semi-arid climates (Thorne 1972; Adams 2004; Farjon 2005; Adams 2008; Mao et al. 2010).

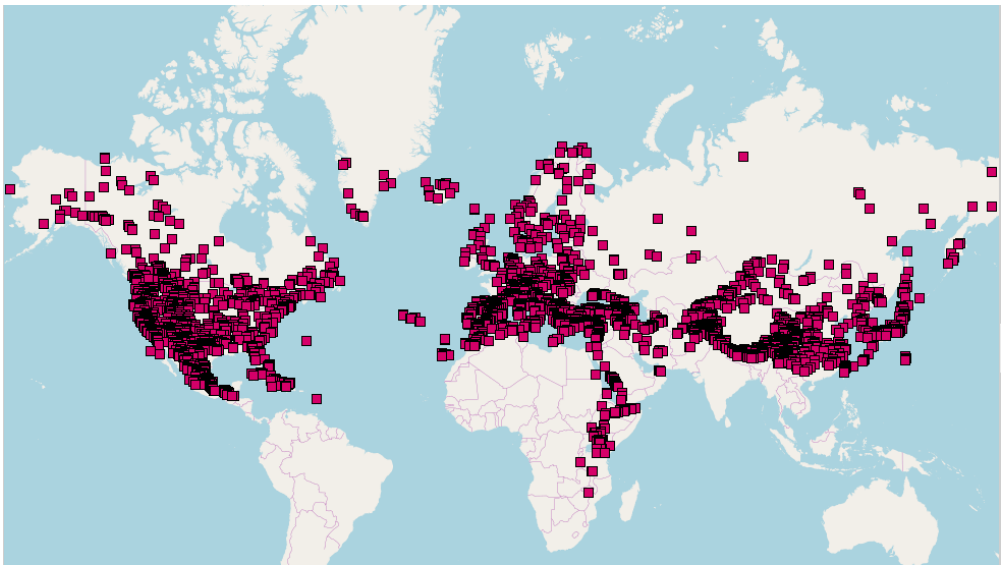

Figure 1. Distribution of *Juniperus* species, generated from herbaria collections in the conifer database: https://herbaria.plants.ox.ac.uk/bol/conifers.

Junipers are generally small, often many-branched evergreen trees or shrubs with short awl, or scale-like leaves, similar to other members of Cupressaceae. *Juniperus* species exhibit monoecy, dioecy, or both reproductive morphologies (Schupp et al. 1997; Teixeira et al. 2014; Farjon 2005); they are wind pollinated and reproduce via seeds in fleshy, berry-like female cones. These seed cones are indehiscent, a unique characteristic among conifers, and a defining character used to distinguish *Juniperus* from closely related genera within the family Cupressaceae (Little 2006). Passerine birds, specifically thrushes, and small mammals have been identified as the primary dispersal agents of *Juniperus* seeds (Livingston 1972; Zamora 1990; Jordano 1993; Schupp et al. 1997; Santos et al. 1999; García 2001; Rumeu et al. 2011). Dispersal of juniper seeds across long distances by birds has likely facilitated the wide distribution of the genus and may explain its occurrence on landmasses that are currently, or were previously, separated by large bodies of water (Mao et al. 2010).

*Juniperus* species are utilized for a few notable commercial products. Juniper berries, primarily from the species *J. communis*, are the essential ingredient and flavoring agent in the production of gin. Juniper wood is highly resistant to termites and fungal decomposition, (Morrell 2010) and commonly used for fence posts (Morrell 1999) and as an exterior building material in the western United States. Juniper wood, as well as wood from other genera in the Cupressaceae family, is also used in the production of "cedarwood oil," a popular fragrance in cosmetic and household products (Adams 2014). In 1984, the United States was the leading producer of cedarwood oil (Lawrence 1985; Adams 2014). Juniper plants are also popular in the horticultural field for use as landscape plants, with 220 known cultivars (Krussmann 1991).

Carl Linneaus (1754) attempted the earliest classification of *Juniperus* species in 1754 (Farjon 2005), with several naturalists and taxonomists struggling to classify *Juniperus* species since then. Among-species variation is limited in this group, and taxonomically-informative morphological features are few. Some characters utilized in early *Juniperus* classification include the size and shape of the seed cone, leaf morphology, and number of seeds per seed cone (Farjon 2005). Early classification of the genus focused on grouping large sections.

Edouard Spach (1841), a French botanist, used foliar characteristics to divide members of the genus into the two sections *Oxycedrus* (=*Juniperus*) and *Sabina*. The Austrian botanist Stephan Endlicher added the third section, *Caryocedrus* in 1847 (Farjon 2005), based on seed fusion and cone scale ridges. This division of *Juniperus* into the 3 major sections: *Juniperus, Sabina,* and *Caryocedrus,* is consistent with modern groupings (Adams and Schwarzbach 2013b; Mao 2010). Currently, section (sect.) *Caryocedrus* contains only one species, *J. drupacea*, located in the Mediterranean. Sect. *Juniperus* is the second largest section and contains 14 species: 12 eastern hemisphere species; one species restricted to western North America; and, one species, the most widely distributed conifer, *J. communis*, distributed across the globe at northern latitudes (Echenwalder 2009). Sect. *Sabina* is the largest of the three sections and contains approximately 60 species, roughly eighty percent of all *Juniperus* species. This section is widely distributed across Europe, Asia, and North America, and is also the most variable in seed cone color and seed number.

Although classification of *Juniperus* into three major sections was achieved relatively early (1847), little advancement in understanding of the relationships among

groups within these sections was gained over the next 150+ years. A continued search for characters, from which to infer relatedness, included investigations into morphology of male and female cones (Lemoine-Sebastian 1967, 1968), the chemical constituents of volatile oils, as well as Random Amplified Polymorphic DNA (RAPD) (Adams and Demeke 1993; Adams 1994; Adams et al. 2006). Despite significant effort and investment, no morphological character, biochemical compound or RAPD data provides consistent and accurate evidence to resolve inter-specific relationships (Farjon 2005). These results underline the need for a better phylogenetic understanding of relationships within the genus.

Two groups of researchers have recently used DNA sequence data to investigate the phylogenetic relationships within *Juniperus*. Mao et al. 2010, using >10,000 bp of chloroplast DNA (cpDNA) from 51 *Juniperus* species, produced the first large-scale phylogeny of the genus, including members of all major clades. Adams and Schwarzbach performed several preliminary phylogenetic investigations into groups of *Juniperus* taxa (2012a, 2012b, 2012c, 2013a, 2013b), with a culminating phylogeny (2013b) including all recognized species of *Juniperus* (~75). Their work was based on one nuclear-ribosomal DNA (nrDNA) region (ITS), and four cpDNA regions, totaling 4411 bp.

Although these two phylogenies agree on the division of *Juniperus* into the three sections *Caryocedrus*, *Juniperus*, and *Sabina*, they do not entirely agree on the circumscription of, or relationships among, groups within sect. *Sabina*. Mao et al. (2010) divides the section into five groups, the monophyletic clades supported in their phylogeny, which they identify with Roman numerals (I-V) (Figure 2A). In contrast,

Adams and Schwarzbach (2013b) identify three major monophyletic groups described by morphology and location: the "serrate-leaf junipers of North America," the "turbinate-seed cone, single-seeded, entire-leaf junipers," and the "multi-seeded, entire-leaf junipers." Groups I and II in Mao et al. (2010) correspond to the turbinate-seed code, single seeded, entire leaf junipers and the serrate-leaf junipers of North America respectively. Clades defined as groups III and IV (Mao et al. 2010) together form the multi-seeded, entire-leaf junipers of Adams and Schwarzbach (2013b). Groups III and IV are each monophyletic clades in the Mao phylogeny, but are intermixed in Adams and Schwarzbach 2013b (Figure 2C), although the placement of the group IV *chinensis* group within the group III clade has low support (posterior probability (0.52)).

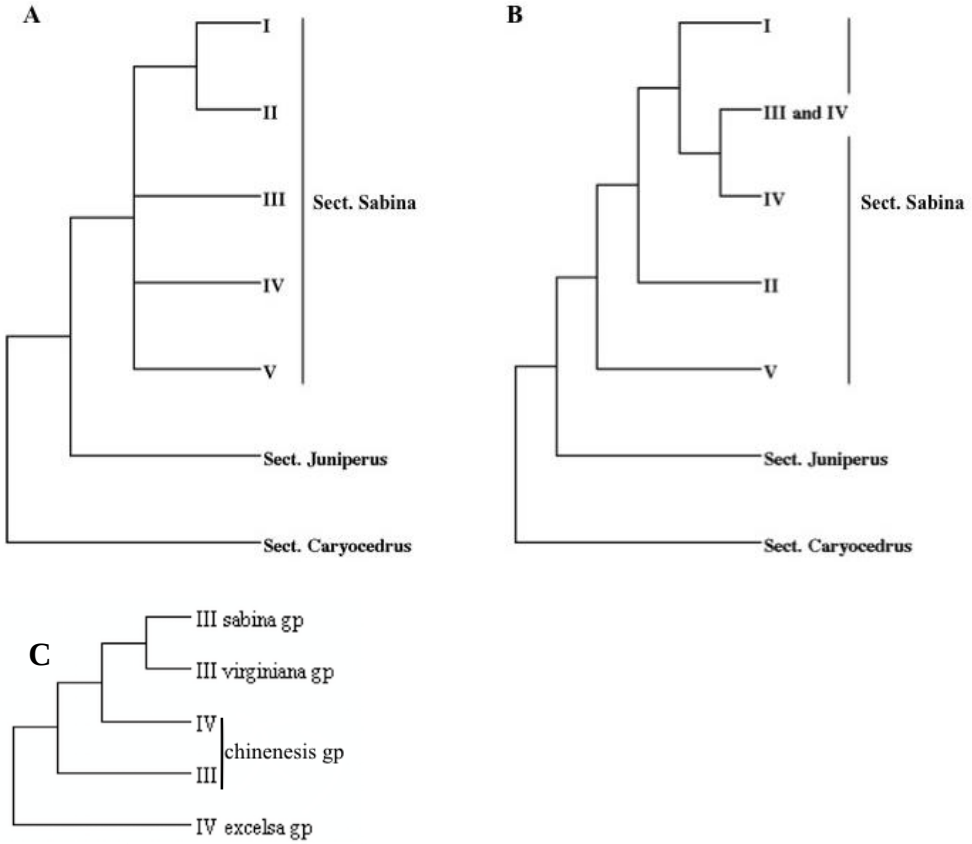


Figure 2. Current *Juniperus* phylogenetic hypotheses simplified to include sections and their major groups. (A) Modified from Mao et al. (2010). (B) Modified from Adams and Schwarzbach 2013b and organized by groups defined in Mao et al. (2010). (C) A subtree of tree B illustrating subgroups of group III and IV, sect. *Sabina* identified in Adams and Schwarzbach (2013b).

It is apparent from the disagreement in topology and lack of resolution or weak support in some areas of the trees that further work can be done to improve our phylogenetic understanding of the genus. A well-resolved, strongly supported topology from the chloroplast genome would provide a solid foundation from which biogeographic and molecular dating analyses could be performed. Although some evidence suggests *Juniperus* originated in Europe, the ancestral area reconstruction analysis of Mao et al. (2010) identified Europe, Asia, or both geographic areas as the geographic area of origin. Further research could narrow the origin location of *Juniperus,* as well as improve our understanding of the timing and mechanism of dispersal events between the eastern and western hemispheres, and the timing or potential causes of divergence.

Although the phylogeny of Adams and Schwarzbach (2013b) is mostly well-resolved, it is generated from a concatenated matrix of chloroplast and nrDNA. Concatenating sequences is effective for DNA with the same evolutionary history, as in the genes of the chloroplast (Yang et al. 2012; Doyle 1992), but can be misleading if the data sets have conflicting signals from differing histories. Mao et al. (2010) tested the whether the chloroplast and nrITS DNA of 24 *Juniperus* species had conflicting phylogenetic signal using the partition homogeneity test. Results of this test indicated "significant contradiction and incongruence" between the nrITS and cpDNA data sets ($p=0.01$).

Incongruent signal from nuclear and chloroplast DNA may be indicative of chloroplast capture by recent or ancient hybridization (Rieseberg and Soltis 1991). Evidence of hybridization in *Juniperus* has been documented for over seventy years (Fasset 1944a,1944b, 1945a, 1945b, 1945c, Hall 1952). Recently, Adams (2015) and

Adams et al. ( 2016, 2017) have investigated the conflicting signal of chloroplast, (petN-psbM), nuclear-ribosomal (nrITS), and a single-copy nuclear gene (maldehy) sequences in closely related pairs or groups of species, revealing evidence of hybridization, chloroplast capture, and introgression. Adams (2015) identified a range of "intermediate" populations extending from eastern Washington and northeast Oregon to western Montana between pure populations of *J. maritima* and *J. scopulorum*. These populations contained individuals with hybrid nrITS, maldehy, or both and possessed the *J. scopulorum* chloroplast genome. Analysis of sequence data from different sources independently, reveals a more accurate and complete story of relatedness, which would be obscured if sequences were concatenated.

The purpose of this study is to improve the current understanding of deep relationships within the genus *Juniperus* by conducting a thorough phylogenetic investigation of the chloroplast genome. This study employs complete and nearly complete chloroplast genomes, > 10x the sequence data used in Mao et al. 2010, and slightly more than one-third of all recognized species. Attention will be focused on resolving among- group relationships within sect. *Sabina*, as these backbone nodes are in conflict or lack resolution in the current juniper phylogenies. Although we employ fewer species than either of the previous phylogenetic works on *Juniperus* (Table 1), the results of these works informed our sampling strategy, designed to address the phylogenetic areas of uncertainty. The taxa are relatively evenly sampled across the genus, with all major groups represented, and taxa sampled from regions of conflict between the two current phylogenetic hypotheses. This study is limited to the scope of a chloroplast

phylogeny, which may not accurately represent the organismal phylogeny, but can be incorporated into research utilizing multiple sources of DNA

Table 1. Comparison of number of species (not including varieties) sampled from the major sections of *Juniperus* in current studies. Group assignment determined by Mao et al. (2010).

| Sections | Balkenbush | Mao et al. 2010 | Adams and Schwarzbach. 2013b |
|---|---|---|---|
| *Caryocedrus* | 1 | 1 | 1 |
| *Juniperus* | 7 | 6 | 14 |
| *Sabina* | 20 | 44 | 60 |
| *Sabina* by Group | | | |
| I | 4 | 10 | 16 |
| II | 5 | 15 | 22 |
| III | 6 | 11 | 21 (Combined III +IV) |
| IV | 4 | 7 | |
| V | 1 | 1 | 1 |
| Total species | 28 | 51 | 75 |

# CHAPTER II

# MATERIALS AND METHODS

## Plant material and overview of samples

This study includes 32 juniper taxa, and five outgroup taxa. Sequences from eight juniper taxa were newly generated for the study, while raw sequence data from 17 taxa were produced previously in the Raubeson lab for a higher-level gymnosperm phylogeny, but remained unanalyzed. Genbank accessions provide an additional seven juniper taxa and five outgroup species, which represent members of the sister clade to *Juniperus* in Cupressaceae (Table 2).

Two library preparation methods, PCR-based and total-genomic, were employed at different times to prepare the two groups of *Juniperus* samples for sequencing. Seventeen PCR-based libraries were prepared, previously, from total genomic DNA contributed by collaborators. Eight total-genomic libraries were prepared for the current study from genomic DNA extracted from fresh leaf tissue. The fresh tissue samples used in total-genomic libraries were received from the Montgomery Botanical Center, the Denver Botanic Gardens, and the Arnold Arboretum (Table 2). These two library types were sequenced several years apart by platforms at different stages of sequencing technology. PCR-based libraries were sequenced earlier and read pools required quality control and filtering. Total-genomic libraries were sequenced recently and did not require filtering of raw reads. All chloroplast genomes were constructed from read pools by a process that combines *de novo* and reference-guided assembly.

Table 2. Taxa used in current study including accession and voucher information. AA= Arnold Arboretum at Harvard, MBC= Missouri Botanical Center, and DBG= Denver Botanic Gardens.

| Taxon | Collaborator code | Genbank accession ID | Accession no. | Source | Voucher no. |
|---|---|---|---|---|---|
| ***Juniperus*** | | | | | |
| Section *Caryocedrus* | | | | | |
| *J. drupacea* | JU8795 | | | | Adams 8795 |
| Section *Juniperus* | | | | | |
| *J. cedrus* | | NC_028190 | | | JCED20150806 |
| *J. communis* | | NC_035068 | | | |
| *J. communis v. depressa* | JU8572 | | | | Adams 8572 |
| *J. formosana* | | KX832625 | | | Yi16054 |
| *J. formosana v. formosana* | JU9071 | | | | Adams 9071 |
| *J. navicularis* | JU8240 | | | | Adams 8240 |
| *J. oxycedrus* | JU9056 | | | | Adams 9056 |
| *J. rigida* | | | 870667*1 | DBG | |
| Section *Sabina* | | | | | |
| *J. arizonica* | JUAR | | | | Adams 7635 |
| *J. ashei v. ashei* | JU6746 | | | | Adams 6747 |
| *J. barbadensis v. lucayana* | | | 20090472*A | MBC | |
| *J. bermudiana* | | NC_024021 | | | Adams 11080 |
| *J. brevifolia* | | | 20080523*A | MBC | |
| *J. californica* | JU10148 | | | | Adams 10148 |
| *J. chinensis* | | | 132406*1 | DBG | |
| *J. chinensis v. procumbens* | | | 791792*1 | DBG | |
| *J. deppeana v. deppeana* | JU9056 | | | | Adams 9056 |
| *J. erectopatens* | JU8532 | | | | Adams 8532 |
| *J. excelsa* | | | 17-2003*B | AA | |
| *J. monosperma* | | NC_024022 | | | Adams 13595 |
| *J. morrisonicola* | JU8681 | | | | Adams 8681 |
| *J. phoenicea* | JU7078 | | | | Adams 7078 |
| *J. pingii* | JU8506 | | | | Adams 8506 |
| *J. procera* | | | 20090473*A | MBC | |
| *J. sabina* | JU7836 | | | | Adams 7839 |
| *J. scopulorum* | | NC_024023 | | | Adams 13594 |
| *J. squamata* | | | 802160*1 | DBG | |
| *J. tibetica* | JU8516 | | | | Adams 8516 |
| *J. virginiana* | | NC_024024 | | | Adams 13549 |
| *J. virginiana v. silicicola* | JU2775 | | | | Adams 2775 |
| *J. virginiana v. virginiana* | JU6746 | | | | Adams 6746 |
| ***Hesperocyparis*** | | | | | |
| *H. glabra* | | | | | Adams 9386 |
| ***Cupressus*** | | | | | |
| *C. chengiana* | | | | | Adams 8130 |
| *C. gigantea* | | | | | Adams 8270 |
| *C. nootkatensis* | | | | | Adams 9403 |
| *C. sempervirens* | | | | | Adams 9216 |

## PCR-based Libraries

### *Library preparation and sequencing*

Chloroplast genomes were PCR amplified from total-genomic DNA using a series of overlapping primer pairs. The sequencing strategy followed Solexa sequencing-by-synthesis procedures (Cronn et al. 2008). An unpublished *Cryptomeria japonica* chloroplast genome was used as the reference genome for the overlapping primer design strategy. Solexa libraries were prepared from these PCR products. The PCR products were pooled and sheared, with unique adapters ligated to the fragments of each library. The fragmented library was size selected and enriched by PCR. The libraries were multiplexed and sequenced on an Illumina platform at Oregon State University.

### *DNA quality control and filtering*

I downloaded the raw sequence data from the PCR-based libraries from a server at Oregon State University. Each file contained a (56 bp) read pool of between 188,022 and 3,459,440 reads, (Table 3) corresponding to a single *Juniperus* taxa. Barcodes (4 bp) and adapters were removed from the reads and they were trimmed and filtered using Trimmomatic (Bolger et al. 2014). The 3' and 5' ends were trimmed to remove bases with a Phred score of 3 or below (Bolger et al. 2014). The function SLIDINGWINDOW:4:15 was employed to further filter *J. virginiana v. virginiana* reads. Reads less than 25 bp were removed from each read pool. Contamination by other gymnosperm DNA was detected in some read pools. The BBsplit tool from the BBMap package version 37.72 (Bushnell) was used to remove contaminating reads, by separating reads that preferentially matched non-*Juniperus* taxa.

Table 3. *Juniperus* taxa sequenced for current study, including reference genome, library preparation method, and number of reads generated.

| Taxon | Reference | Ilumina library | Raw read no. |
|---|---|---|---|
| *J. arizonica* | *J. monosperma* | PCR | 1896828 |
| *J. ashei v. ashei* | *J. monosperma* | PCR | 1856219 |
| *J. barbadensis* | *J. bermudiana* | Total Genomic | 62090378 |
| *J. brevifolia* | *J. cedrus* | Total Genomic | 54604789 |
| *J. californica* | *J. monosperma* | PCR | 638219 |
| *J. chinensis* | *J. bermudiana* | Total Genomic | 72545494 |
| *J. communis v. depressa* | *J. communis* | PCR | 769143 |
| *J. deppeana v. deppeana* | *J. monosperma* | PCR | 531083 |
| *J. drupacea* | *J. cedrus* | PCR | 391177 |
| *J. erectopatens* | *J. bermudiana* | PCR | 2387201 |
| *J. excelsa* | *J. bermudiana* | Total Genomic | 88378821 |
| *J. formosana v. formosana* | *J. formosana* | PCR | 1068707 |
| *J. morrisonicola* | *J. monosperma* | PCR | 836922 |
| *J. navicularis* | *J. cedrus* | PCR | 112806 |
| *J. oxycedrus* | *J. cedrus* | PCR | 2194783 |
| *J. phoenicea* | *J. monosperma* | PCR | 440573 |
| *J. pingii* | *J. monosperma* | PCR | 490279 |
| *J. procera* | *J. bermudiana* | Total Genomic | 65199513 |
| *J. procumbens* | *J. bermudiana* | Total Genomic | 57130084 |
| *J. rigida* | *J. communis* | Total Genomic | 60376494 |
| *J. sabina* | *J. bermudiana* | PCR | 1429407 |
| *J. squamata* | *J. monosperma* | Total Genomic | 55429535 |
| *J. tibetica* | *J. monosperma* | PCR | 2742708 |
| *J. virginiana v. silicicola* | *J. virginiana* | PCR | 3459440 |
| *J. virginiana v. virginiana* | *J. virginiana* | PCR | 188022 |

## Total genomic libraries

### *DNA extraction*

Leaf-tissue samples were ground into liquid nitrogen and maintained at -80 °C. Total genomic DNA was extracted using a Nucleon PhytoPure Genomic DNA extraction Kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK). A modified cetyltrimethylammonium bromide (CTAB) extraction procedure (Doyle and Doyle 1987) with an additional chloroform: isoamyl alcohol extraction step was used for two samples

that were not cleanly extracted by the PhytoPure kit. Concentrations of DNA extractions were quantified using the Qubit dsDNA BS Assay Kit and Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, California, USA).

*Library preparation and sequencing*

Libraries were prepared using the NxSeq AmpFree Low DNA Library Kit with Adaptors (Lucigen Corporation, Middletown, WI, USA), and quantitated using the same procedures as above. The libraries were size selected and sequenced at the Genomics Core facility, WSU Spokane. The eight libraries were multiplexed and sequenced on the Illumina HiSeq2500 platform to generate 400 million 100 bp, single-end reads. Pre-processing, including trimming of barcodes and Illumina adapters, and removal of reads shorter than 35 bp, was performed at the sequencing facility. Raw reads had 100% high quality when examined in Geneious, so quality trimming of the total-genomic libraries was deemed unnecessary.

**Plastome assembly**

Reference genomes of seven *Juniperus* species were downloaded from Genbank (Table 1). Reference-guided assembly of each read pool was performed using the most closely-related reference genome available. Assemblies were implemented in Geneious v.8.1.9 (Biomatters Ltd., Auckland, NZ) with medium-low sensitivity and up to 25 iterations for taxa from PCR-based libraries, and medium-low sensitivity with up to 5 iterations for taxa from total-genomic libraries. Consensus sequences were generated under the highest-quality threshold and positions with less than 5x coverage masked by gaps (Straub et al. 2012). Sequences were aligned using MAFFT version 7 (Katoh &

Standley 2013). The final alignment of the complete chloroplast genomes contained 82 protein-coding genes, 4 rRNAs, and 33 tRNAs.

**DNA sequencing and quality control filtering**

Sequencing of PCR-based libraries generated read pools of between 112,806 and 3,459,440 reads with an average of 1,260,795 reads (Table 4). Quality control and filtering reduced PCR-based read pools by between 95.43 and 6.15 percent, resulting in an average of 794,746 reads retained. Sequencing of total-genomic libraries produced between 54,604,789 and 88,378,821 reads, with an average of 64,469,389 reads.

Table 4. Sequencing results including raw read number and number of reads retained following quality control and filtering. Numbers in bold indicated final read number used for plastome assembly.

| Taxon | Raw read | Post-trim | Reads | Post- | Reads | Raw reads |
|---|---|---|---|---|---|---|
| *J. arizonica* | 1896828 | 148190 | 7.81 | **86671** | 58.49 | 4.57 |
| *J. ashei v. ashei* | 1856219 | **1742099** | 93.85 | - | - | 93.85 |
| *J. barbadensis v. lucayana* | **62090378** | - | - | - | - | 100 |
| *J. brevifolia* | **54604789** | - | - | - | - | 100 |
| *J. californica* | 638219 | 528140 | 82.75 | **457082** | 86.55 | 71.62 |
| *J. chinensis* | **72545494** | - | - | - | - | 100 |
| *J. communis v. depressa* | 769143 | 704923 | 91.65 | **555099** | 78.75 | 72.17 |
| *J. deppeana v. deppeana* | 531083 | 407058 | 76.65 | **234635** | 57.64 | 44.18 |
| *J. drupacea* | 391177 | **376694** | 96.3 | - | - | 96.30 |
| *J. erectopatens* | 2387201 | 2251685 | 94.32 | **1967746** | 87.39 | 82.43 |
| *J. excelsa* | **88378821** | - | - | - | - | 100 |
| *J. formosana v. formosana* | 1068707 | 891104 | 83.38 | **682715** | 76.61 | 63.88 |
| *J. morrisonicola* | 836922 | 447694 | 53.49 | **290920** | 64.98 | 34.76 |
| *J. navicularis* | 112806 | 71059 | 62.99 | **27198** | 38.28 | 24.11 |
| *J. oxycedrus* | 2194783 | 789915 | 35.99 | **626571** | 79.32 | 28.55 |
| *J. phoenicea* | 440573 | 400297 | 90.86 | **277081** | 69.22 | 62.89 |
| *J. pingii* | 490279 | 386842 | 78.9 | **260351** | 67.30 | 53.10 |
| *J. procera* | **65199513** | - | - | - | - | 100 |
| *J. procumbens* | **57130084** | - | - | - | - | 100 |
| *J. rigida* | **60376494** | - | - | - | - | 100 |
| *J. sabina* | 1429407 | 1362238 | 95.3 | **1174627** | 86.23 | 82.18 |
| *J. squamata* | **55429535** | - | - | - | - | 100 |
| *J. tibetica* | 2742708 | 2618754 | 95.48 | **2195287** | 83.83 | 80.04 |
| *J. virginiana v. silicicola* | 3459440 | 2974659 | 85.99 | **2456187** | 82.57 | 71.00 |
| *J. virginiana v. virginiana* | 188022 | **99726** | 53.04 | - | - | 53.04 |

**Plastome assembly**

The coverage of the plastome is an average of the number of reads that contain each nucleotide position. The average coverage of all taxa was 340, with an average standard deviation (SD) of 326. The samples prepared using PCR and total-genomic libraries had very similar average coverage (Table 5)), with means of 334.94 and 350.25, but taxa prepared with PCR-based libraries exhibited a significantly higher variance of read depth (F-test, p=0.0009).

Table 5. Plastome assembly details including number of mapped reads to reference, coverage of the reference genome, the length of the original sequence and length of the sequence with masking of regions with less than 5x coverage. HQ bases refer to a Phred quality score ≥ Q40.

| Taxon | Reads mapped to ref. (#) | Ave. coverage ± 1 SD | Assembled length | % HQ bases | Masked length | %HQ bases |
|---|---|---|---|---|---|---|
| *J. arizonica* | 86573 | 37 ±50 | 125909 | 91.3 | 108877 | 99.9 |
| *J. ashei v. ashei* | 1450927 | 633 ±921 | 104437 | 78.7 | 98844 | 100 |
| *J. barbadensis v. lucayana* | 446668 | 348 ±38 | 127641 | 100 | - | - |
| *J. brevifolia* | 590893 | 462 ±57 | 127144 | 100 | - | - |
| *J. californica* | 456561 | 199 ±360 | 127195 | 99.1 | 123665 | 100 |
| *J. chinensis* | 417347 | 325 ±48 | 127758 | 100 | - | - |
| *J. communis v. depressa* | 551565 | 239 ±221 | 127141 | 84 | 104772 | 100 |
| *J. deppeana v. deppeana* | 625853 | 272 ±448 | 123599 | 91 | 110564 | 100 |
| *J. drupacea* | 284546 | 126 ±105 | 117976 | 94.1 | 117132 | 100 |
| *J. erectopatens* | 1963953 | 850 ±1063 | 123498 | 87.3 | 106911 | 100 |
| *J. excelsaexcels* | 579508 | 450 ±63 | 127769 | 100 | - | - |
| *J. formosana v. formosana* | 678640 | 295 ±326 | 123209 | 82.7 | 97825 | 100 |
| *J. morrisonicola* | 290326 | 126 ±259 | 125723 | 81.2 | 82769 | 99.9 |
| *J. navicularis* | 27091 | 12 ±16.7 | 125494 | 80.1 | 80133 | 99.9 |
| *J. oxycedrus* | 234451 | 103 ±113 | 124398 | 85.9 | 104118 | 100 |
| *J. phoenicea* | 276863 | 121 ±116 | 125133 | 90.4 | 109435 | 100 |
| *J. pingii* | 260083 | 113 ±181 | 123417 | 87.6 | 106541 | 99.9 |
| *J. procera* | 321402 | 250 ±36 | 127731 | 100 | - | - |
| *J. procumbens* | 278176 | 216 ±31 | 127753 | 100 | - | - |
| *J. rigida* | 561803 | 435 ±68 | 128449 | 99.9 | - | - |
| *J. sabina* | 1172302 | 508 ±818 | 125343 | 87.4 | 108852 | 100 |
| *J. squamata* | 403486 | 316 ±53 | 127682 | 100 | - | - |
| *J. tibetica* | 2191835 | 953 ±1041 | 120796 | 90 | 104092 | 100 |
| *J. virginiana v. silicicola* | 2451970 | 1074 ±1665 | 125248 | 64.7 | 69474 | 99.9 |
| *J. virginiana v. virginiana* | 76462 | 33 ±53 | 90325 | 71.5 | 82380 | 99.9 |

## Phylogenetic Analysis

Three alignments were used for phylogenetic analysis, a complete chloroplast sequence alignment, a protein-coding gene and rRNA alignment, and a protein-coding gene-only alignment. Four partition schemes of the complete alignment were explored: 1. no partition, 2. partitioned by gene, 3. partition by gene and rRNA and 4. partitioned by gene, rRNA, and noncoding region. The gene + rRNA alignment was

evaluated without partitioning and with partitioning by gene and rRNA. The un-partitioned gene-only alignment was explored as well as partitioning by gene. Maximum parsimony (MP) and maximum likelihood (ML) analyses were performed on each alignment and partition scheme (Table 6). Bayesian inference (BI) was performed on the un-partitioned and maximally partitioned gene-only and complete alignments.

*Maximum Parsimony*

MP analysis was performed in PAUP 4.0a152 for Macintosh. Character states were considered unordered and of equal weight. A heuristic search was employed with the starting tree obtained via stepwise addition and one tree held at each step. TBR branch swapping, steepest descent, MulTrees, and Collapse options were in effect. No upper limit was set for the number of trees held in memory. Bootstrapping with 1,000 replicates was conducted.

*Maximum Likelihood*

ML analysis was performed in IQ-TREE version 1.5.5 (Minh et al. 2017). ModelFinder (Kalyaanamoorthy et al. 2017) was employed to select the best-fit models for each alignment and data partition, determined by the corrected Akaike information criterion (AICc). The DNA model of substitution was applied in all analyses, with the edge-proportional (-spp) partition model (Chernomor et al. 2016) in effect. ML trees were constructed under the best-fit partition model (Nguyen et al. 2015), and 1,000 ultrafast (uf) bootstrap (BS) replicates performed (Mihn et al. 2013). It is important to note that the interpretation of uf bootstrap support values is slightly different than that of normal nonparametric bootstrap values. Uf BS values are considered to reflect

probability that the clade is true, and only clades with ≥95 uf BS are considered strongly supported (Mihn et al. 2013).

IQ-TREE was used to infer site-specific evolutionary rates of nucleotides in the un-partitioned complete and gene-only data matrices. Site-specific rates of each nucleotide were written to file using the –wsr command, and each nucleotide was assigned a Gamma rate category. Nucleotide positions were sorted into alignment files according to rate category. Individual rate categories and intuitive combinations of rate categories were evaluated by MP and ML analysis (Table 6).

*Bayesian Inference*

Bayesian inference analyses were performed in MrBayes v.3.2 (Ronquist et al. 2012). Two independent Markov Chain Monte Carlo (MCMC) runs of 1,000,000 generations were performed simultaneously under the GTR+G+I model, with three cold, and one incrementally heated chain. A burn-in fraction was set to 0.25. The analyses were run until the two independent MCMC runs converged, indicated by a standard deviation of split frequencies less than 0.01 (Ronquist et al. 2012). The estimated sample size of each parameter was evaluated to ensure adequate sampling (>100), and the potential scale reduction factor was also checked for nearness to one.

Table 6. Data sets and partitioning schemes evaluated by phylogenetic analysis, and type of analyses performed on each. x=analysis performed - = analysis not performed

| Alignment | Partition scheme | ML | MP | BI |
|---|---|---|---|---|
| **Gene** | | | | |
| | None | x | x | x |
| | Gene | x | x | x |
| rate 1 | - | - | x | - |
| rate 4 | - | x | x | - |
| rate 5 | - | x | x | - |
| rate 4+5 | | x | x | - |
| **Gene+rRNA** | | | | |
| | None | x | x | - |
| | gene + rRNA | x | x | - |
| **Complete** | | | | |
| | None | x | x | x |
| | Gene | x | x | - |
| | gene + rRNA | x | x | - |
| | gene + rRNA + noncoding | x | x | x |
| rate 1 | - | - | - | - |
| rate 3 | - | x | x | - |
| rate 4 | - | x | x | - |
| rate 5 | - | x | x | - |
| rate 3+4 | - | x | x | - |

*Hypothesis testing*

The Approximately Unbiased (AU) test (Shimodaira 2002) was implemented to evaluate the power of the two data matrices to reject alternative relationships among groups in sect. *Sabina*. The test included the three topologies generated from analyses in the current study, as well as the phylogeny presented in Adams and Schwarzbach 2013b, and the only fully resolved topology presented in Mao et al. (2010) (BEAST and Bayes-DIVA analyses). A sixth tree was included in the test, which combined elements of the Adams and Schwarzbach and Mao *et al* trees. The tree contained the relationships among groups in sect. *Sabina* proposed by Adams and Schwarzbach, but the groups

within sect. *Sabina* were circumscribed as in Mao et al. (2010) (monophyletic groups III and IV).

The AU was implemented in PAUP using parameters estimated empirically by IQTREE analyses, including the model, substitution rate matrix, nucleotide frequencies, gamma distribution shape, and number of gamma rate categories. The remaining parameters were left on default setting, 5,000 bootstrap replicates were performed, and the AICc information criterion was employed.

The AU test was also employed to evaluate the sequence data of chloroplast regions used in Mao et al. (2010). This data is publically available on Genbank, and was downloaded for all *Juiperus* taxa and three outgroup species, with the intention of exploring the resolving power of the data. The AU test evaluated whether the data was strong enough to support one topology, while rejecting other topological hypotheses, including those produced by our current study. The test was employed under the model parameters identified by ML analysis in IQTREE. Indel data was not available, so they were not included, as in Mao et al. (2010).

# CHAPTER III

# RESULTS

## Phylogenetic Analyses

Thirty-two *Juniperus* taxa, including 28 unique species, were used in the phylogenetic analyses. Eight complete chloroplast genomes, ranging from approximately 127-128 kb in length, were generated for this study, and 17 nearly complete genomes were assembled from sequence data generated previously for a study of gymnosperms. Seven complete juniper chloroplast genomes were included from Genbank. The complete aligned data set included 141,328 characters, with 13,269 variable characters, of which 7,289 were parsimony-informative. Phylogenetic analyses were also performed on a subset of the data, the gene-only matrix, containing 76,419 characters, of which 5,628 were variable, and 3,198 of those parsimony-informative.

MP, ML, and BI analyses were performed on the complete matrix as well as different subsets of the data (Table 7). Three topologies among groups within section *Sabina* were recovered from different analyses, but a single topology was most commonly resolved (Figure 3). The most common topology resolved the following relationships among groups within section *Sabina*: V((I,II)(III,IV)), and was supported in 22 of 25 analyses. The alternative topology, (((III,IV)II)I), was produced in two analyses and ((I,II)V)(III,IV) was produced in one analysis (Figure 3). At the species level, the only incongruence among different topologies was in the relationship between *J. californica* and *J. deppeana v. deppeana*. In some instances, *J. californica* is placed basally in group II, and in others it is placed sister to *J. deppeana v. deppeana*.

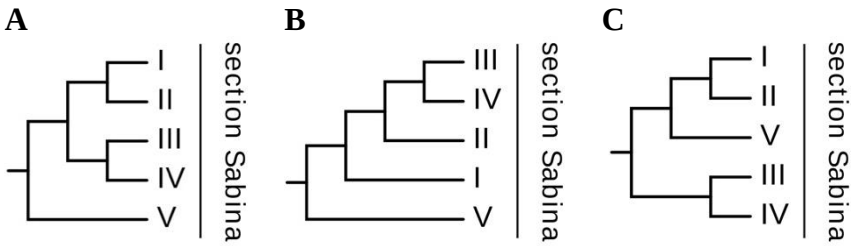

Figure 3. Topologies produced by phylogenetic analyses.

Table 7. Results of phylogenetic analyses of each alignment and partition scheme. (- = not performed)

| Alignment | Partition scheme | ML | MP | BI |
|---|---|---|---|---|
| **Gene** | | | | |
| | None | A | B | A |
| | Gene | A | - | A |
| **Gene+rRNA** | | | | |
| | None | A | B | - |
| | gene + rRNA | A | - | - |
| **Complete** | | | | |
| | None | A | A | C |
| | Gene | A | - | - |
| | gene + rRNA | A | - | - |
| | gene + rRNA + noncoding | A | - | A |

*Maximum Parsimony*

Maximum parsimony analysis of the complete alignment resulted in a single best tree, with a tree score of 17,910 (Figure 4). Section *Sabina* was resolved V((I,II)(III,IV)), although the sister relationship between groups I and II and the placement of groups (I,II) sister to (III,IV) were weakly supported (BS 57, BS 61).

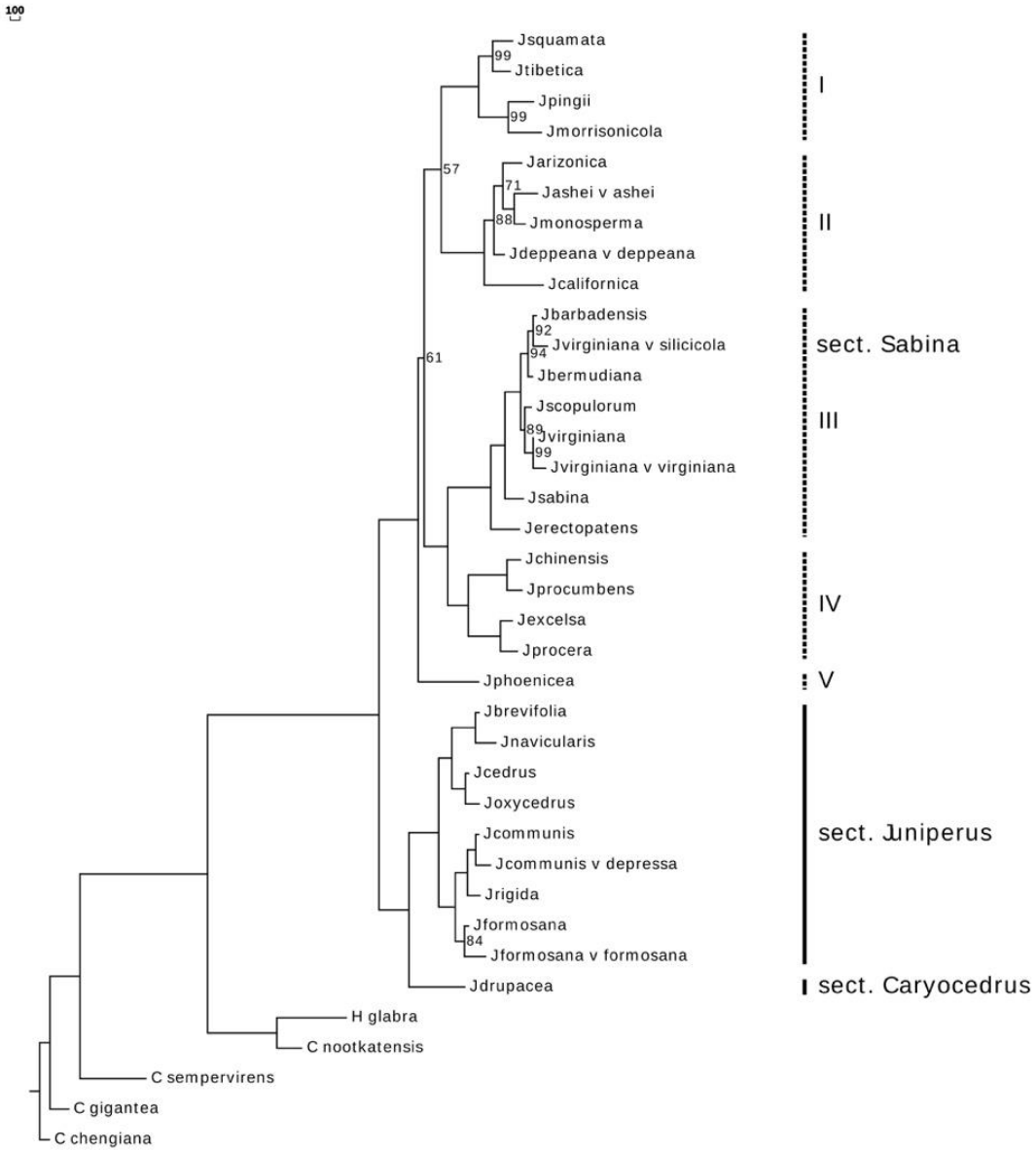


Figure 4. MP tree from the complete alignment with bootstrap support values from 1,000 replicates. All unlabeled nodes have 100 bootstrap support.

MP analysis of the gene-only alignment produced two best trees with a tree score of 7470 (Figure 5). These two trees had a single difference in the relationship between *J. californica* and *J. deppeana v. deppeana.* Analysis of the gene-only alignment produced a different topology of section *Sabina* than the complete alignment. The topology

resolved groups within section *Sabina*, V(I(II(III+IV))), although the branch placing group II sister to (III,IV) was weakly supported (bootstrap 55).

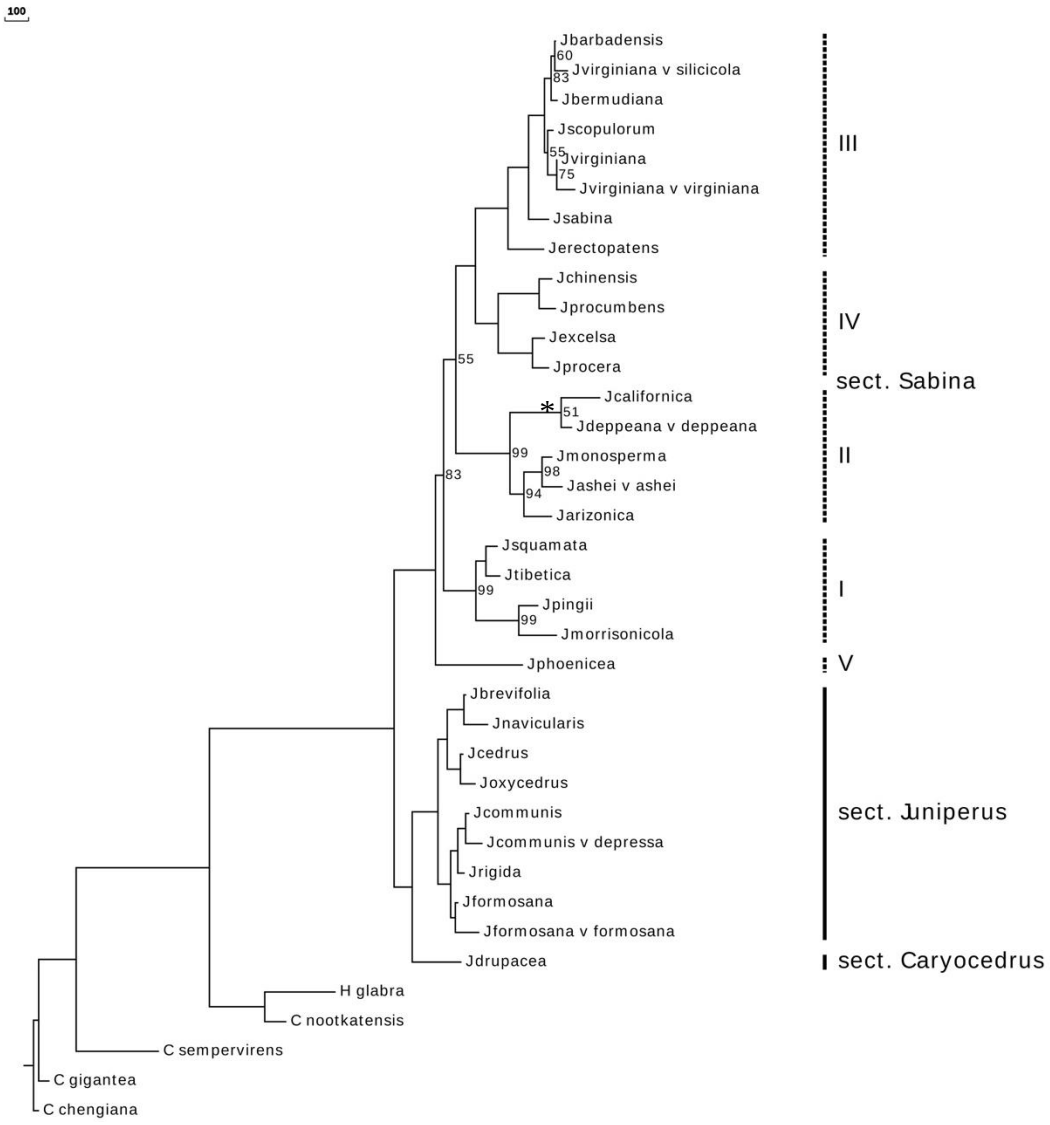


Figure 5. MP tree of the coding-gene alignment with bootstrap support values from 1000 replicates. All unlabeled nodes have 100 bootstrap support. * *J. californica* resolved basally in other MP best-tree.

*Maximum Likelihood*

Maximum likelihood analysis of the complete chloroplast alignment produced one ML tree (Figure 6), identical under all partitioning schemes. The topology of these ML analyses agrees with that of the MP analysis of the complete alignment (Figure 4),

and has improved support for relationships among groups within sect. *Sabina*. While likelihood scores improved with increased partitioning of the complete alignment (Table 8), bootstrap support values did not (Figure 6). The best likelihood score (-lnL = 308132.3) was found by the partitioning of gene, rRNA, and noncoding regions (181 partitions). The –lnL score of the other ML analyses were 315,982.6 (no partitions) and 312,747.6 (coding partition).

Table 8. Summary of maximum likelihood analyses by alignment and partition scheme.

| Alignment | Partition (#) | Alignment length | Informative (%) | Constant (%) | Missing (%) | -lnL (BIC) | -lnL (AICc) |
|---|---|---|---|---|---|---|---|
| Gene only | | 76419 | 4.17 | 92.64 | 4.13 | | |
| | none | | | | | 158193.6 | 158182.0 |
| | gene (82) | | | | | 155541.9 | 155435.8 |
| Gene+ rRNA | | 81139 | 3.98 | 92.92 | 3.89 | | |
| | none | | | | | 166363.7 | 166363.7 |
| | gene and rRNA (86) | | | | | 163314.2 | 163188.9 |
| Complete | | 141328 | 5.14 | 90.69 | 4.44 | | |
| | none | | | | | 316064.0 | 315982.6 |
| | gene (82) | | | | | 312865.5 | 312747.6 |
| | gene and rrna (87) | | | | | 311830.3 | 311701.6 |
| | gene+ rrna+and noncoding (181) | | | | | 308197.2 | 308132.3 |


Figure 6. ML tree of maximally partitioned complete matrix. Uf bootstrap values of complete alignment with all partitions/partitioned by gene/no partition (1000 replicates). All unlabeled nodes have 100 ultrafast bootstrap support. The un-partitioned alignment differed in topology, with a sister relationship between *J. californica* and *J. deppeana v. deppeana*.

ML analysis of the gene-only alignment agreed with ML analysis of the complete alignment, except that *J. californica* and *J. deppeana v. deppeana* are sister taxa in the gene-only alignment, while *J. californica* is the basal taxa of group II in the tree resolved by the partitioned, complete alignment. Partitioning of the gene-only alignment did not change the topology of the ML tree, but it did increase the bootstrap support values of two branches within section *Sabina* (Figure 7). The group (I,II) clade saw increased support from 38 to 66, and the support for groups (I,II) sister to (III,IV) increased from 82 to 90. The remaining branches within the tree received similar support in both analyses. The likelihood scores of the partitioned and un-partitioned gene-only analyses are 155731.45 and 158193.57 respectively.


Figure 7. ML tree from coding-gene alignment partitioned by gene. Bootstrap values from gene alignment are partitioned by gene/no partition and 1000 replicates. All unlabeled nodes have 100 ultrafast BS support.

*Bayesian Inference*

Bayesian analysis of the partitioned and un-partitioned complete alignment produced different topologies of groups within section *Sabina*. Analysis of the partitioned, complete alignment resolved the same best-tree as the MP and ML analyses of the complete alignment (Figure 8), but the un-partitioned, complete alignment resolved a strongly-supported, unique topology, placing group V sister to groups (I,II) (Figure 9).

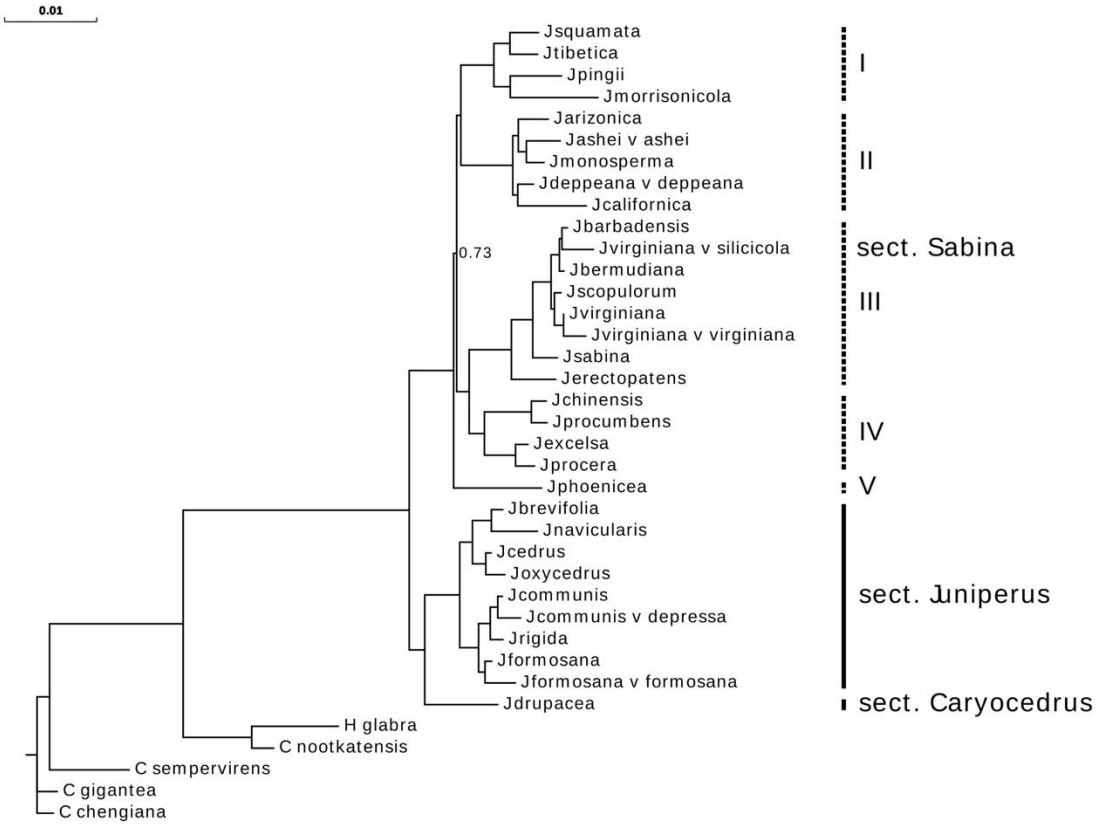


Figure 8. Bayesian inference analysis of the partitioned complete alignment. Posterior probabilities of 1.0 not displayed.

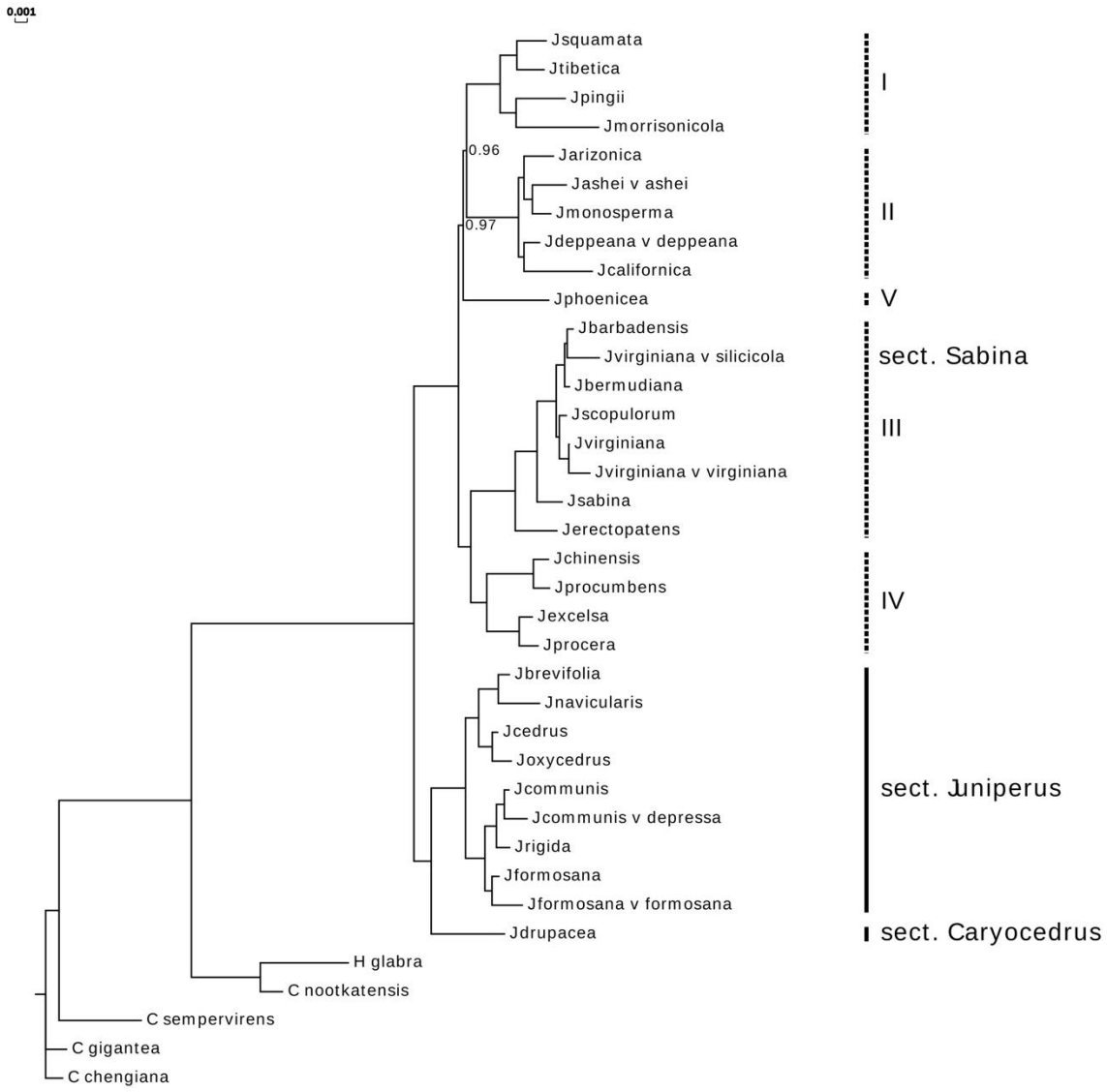


Figure 9. Bayesian inference analysis of the un-partitioned complete alignment. Posterior probabilities of 1.0 not displayed.

Bayesian Inference analysis of the partitioned and un-partitioned gene alignment resolved the same best-tree as BI analysis of the partitioned complete alignment (Figure 10). Partitioning of the gene alignment did not greatly improve posterior probabilities.

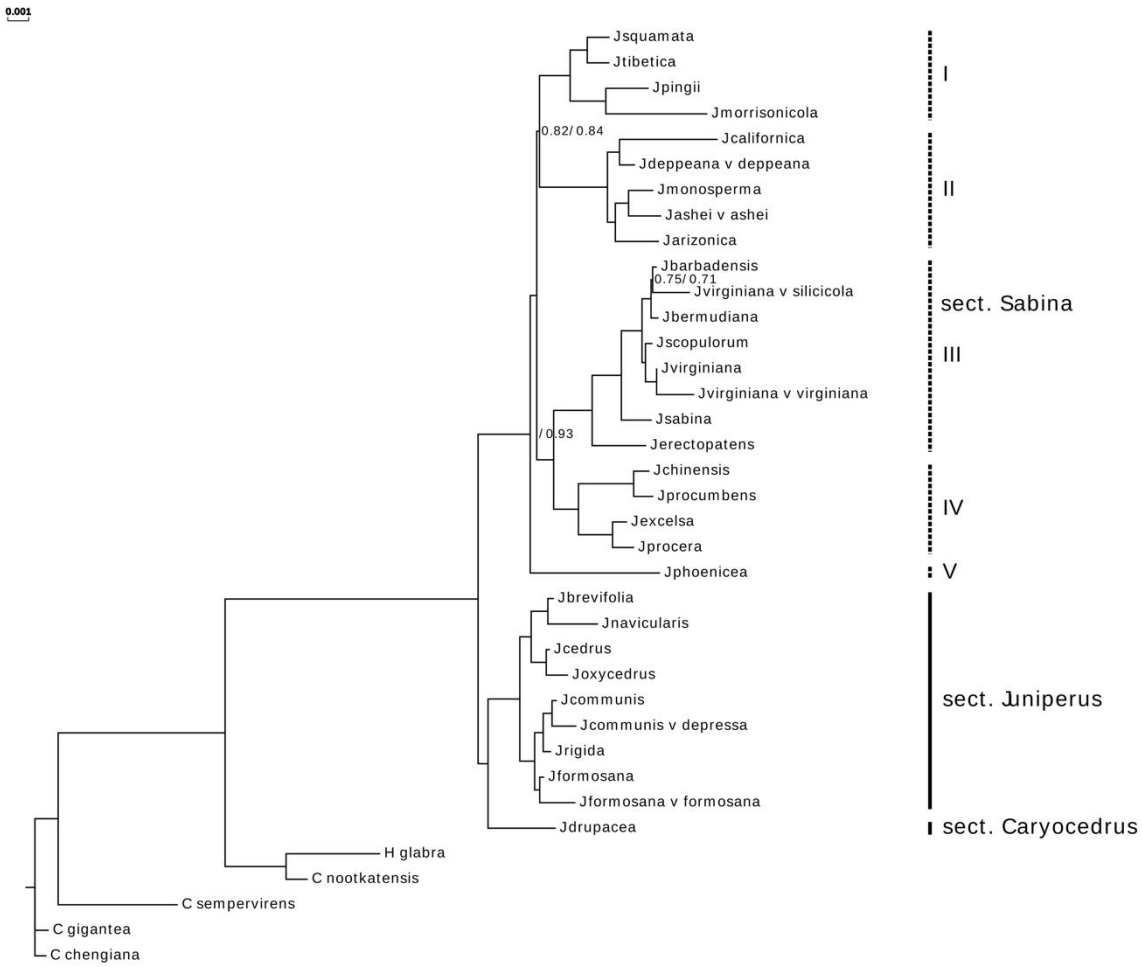


Figure 10. Bayesian analyses of the coding-gene alignment, with posterior probabilities (un-partitioned/partitioned by gene). All nodes with 1.0 posterior probabilities are not labeled.

## Rate Analysis

IQTREE was used to calculate site-specific evolutionary rates of each nucleotide position in the complete and gene matrices during the model-fitting process. Each position was assigned to a gamma distributed rate category, with empirically determined alpha distribution and number of categories. MP and ML analyses were performed on each rate category and intuitive combinations of rate categories. The most effective rate category from each alignment was determined by the number of branches resolved with ≥ 95 BS support.

*Rate analysis-Maximum parsimony*

The rate 3 category of nucleotides was the most informative rate category from the complete alignment, determined by number of highly supported branches, and independent of topology (Table 9). While this category contained only ~4.5% of the data contained in the complete alignment, it resolved the same topology with all but one node supported by ≥ 95 BS. The only difference in topology is the common discrepancy in the relationship between *J. californica* and *J. deppeana v. deppeana* (Figure 11).

Table 9. Summary of MP and ML analyses on nucleotide rate category data sets. Bold indicates most informative rate category from each alignment. MP/ML/BI indicates topologies from Figure 3. U= groups among section *Sabina* not resolved with ≥ 50 BS support. - = analysis not performed. * indicates all positions are variable

| Matrix | Rate Cat. | Positions (#) | Var./var. inform. | Branches resolved | Branches ≥95 ML | Branches ≥95 MP | % Orig. alignment | % PICs | MP/ML /BI |
|---|---|---|---|---|---|---|---|---|---|
| **Complete** | | | | | | | | | |
| | 1 | 128175 | 44/4 | - | - | - | 90.69 | 0.003 | U/U/- |
| | **3** | **6180*** | ***/2765** | **34** | **34** | **33** | **4.37** | **44.74** | **A/A/A** |
| | 4 | 6702* | */4269 | 31 | 25 | 23 | 4.74 | 63.70 | U/U/- |
| | 5 | 271* | */251 | 34 | 5 | 26 | 0.19 | 92.62 | U/U/- |
| | 3+4 | 12882* | */7034 | 34 | 29 | 29 | 9.11 | 54.60 | A/A/- |
| **Genes** | | | | | | | | | |
| | 1 | 70791 | 0/0 | - | - | - | 92.64 | 0.00 | -/-/- |
| | **4** | **5440*** | ***/3013** | **34** | **30** | **29** | **7.12** | **55.39** | **A/A/A** |
| | 5 | 188* | */185 | 19 | 9 | 3 | 0.25 | 98.40 | U/U/- |
| | 4+5 | 5628* | */3198 | 34 | 30 | 24 | 7.36 | 56.82 | A/A/- |

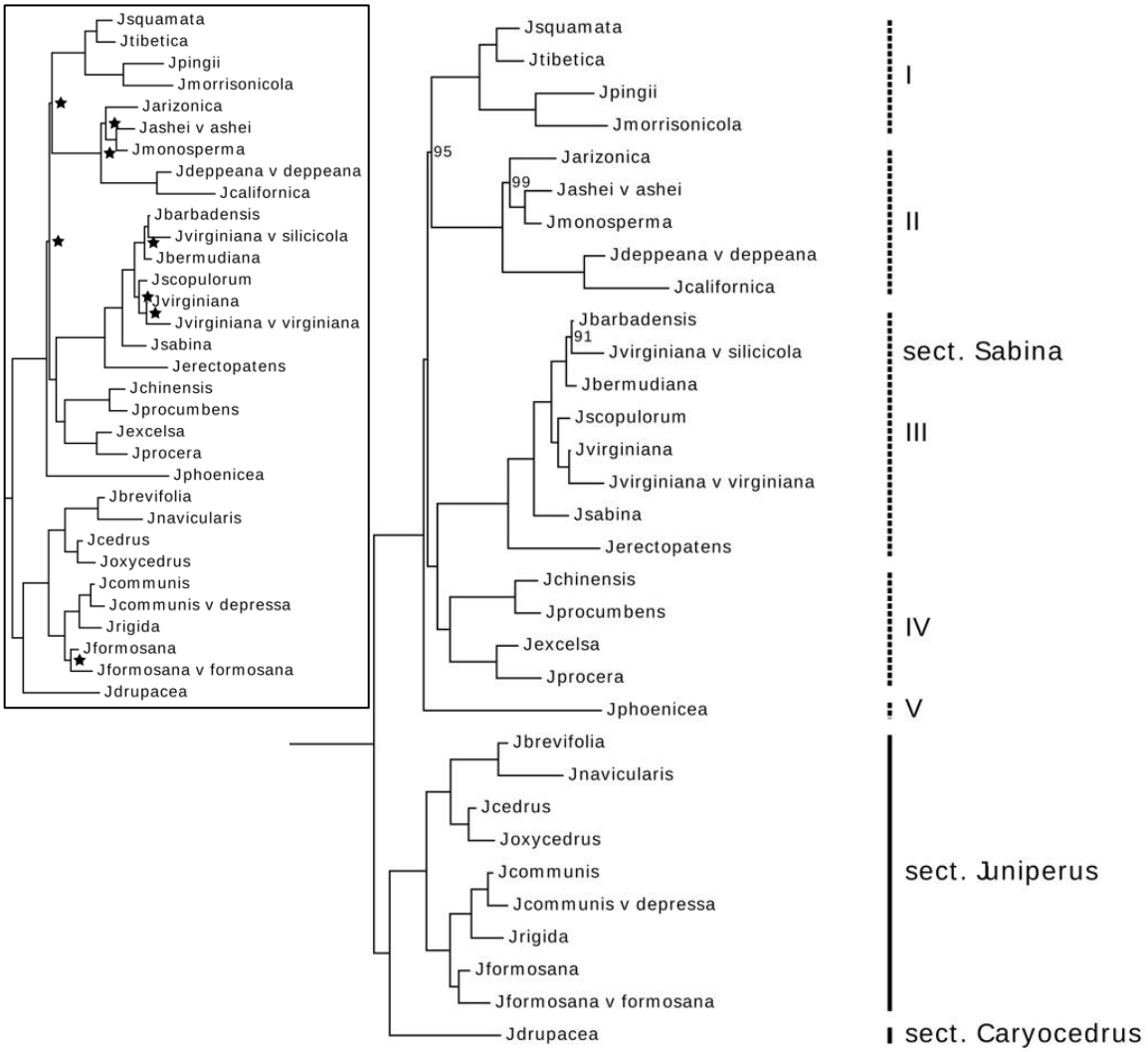


Figure 11. Maximum parsimony tree from complete-rate3 alignment with BS support values from 10,000 replicates. All unlabeled nodes have 100 bootstrap support. Inset: stars indicate increased BS support over MP analysis of complete alignment (Figure 4).

The most informative rate category from the gene alignment was rate 4 (Table 9). The topology produced by this category of nucleotides was the same as the MP analysis of the gene alignment ((((III,IV)II)I)V), but the sister relationship of group II to (III,IV) was weakly supported (BS 57) (Figure 12). In rate analysis of both the complete and gene alignment, the best rate category was also the lowest of the variable rates in each alignment, as rate 1 is invariable or nearly so.

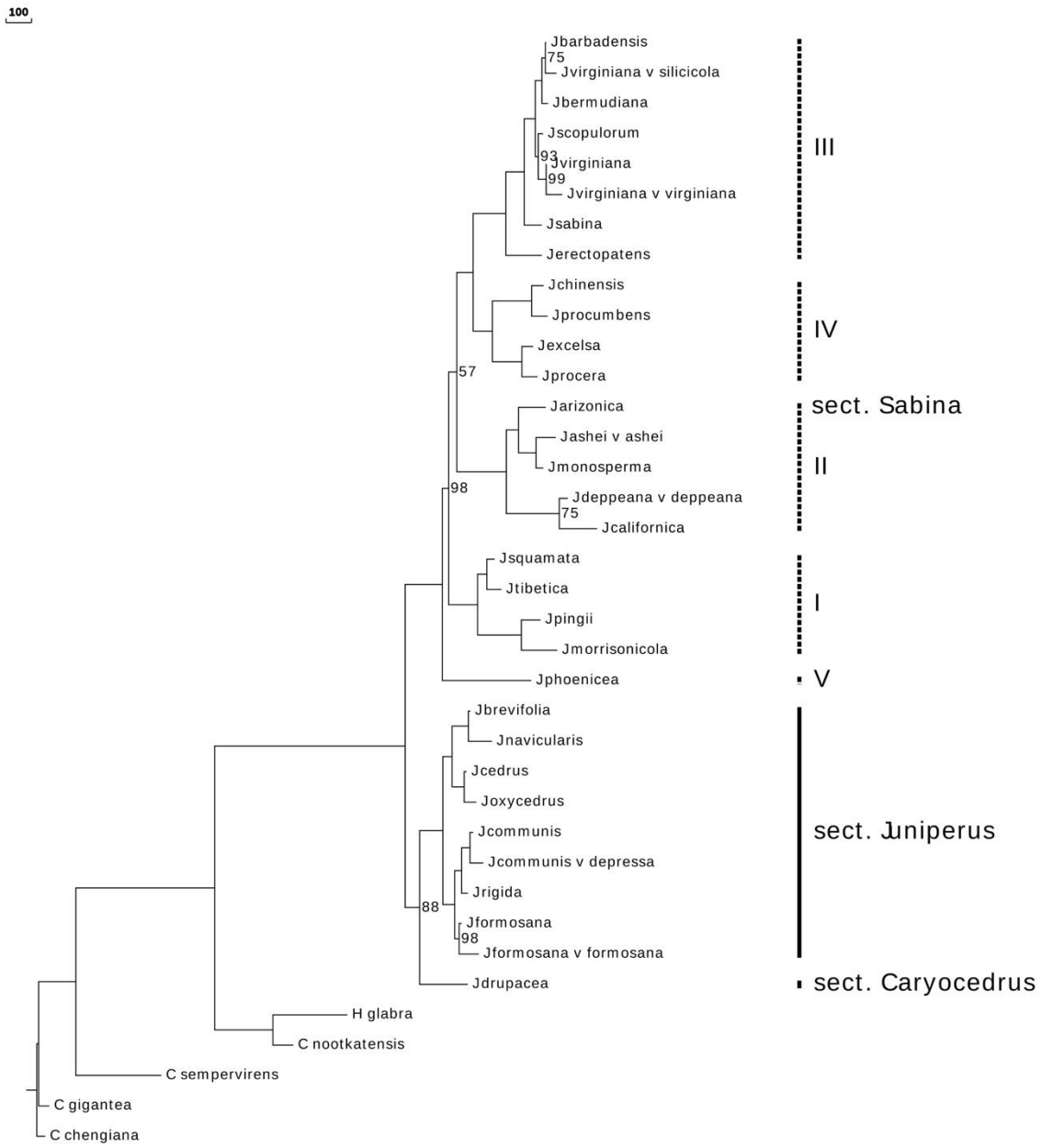


Figure 12. Maximum parsimony tree from gene-rate4 alignment with BS support values from 10,000 replicates. All unlabeled nodes have 100 bootstrap support.

*Rate analysis ML*

ML analysis of nucleotides from rate category 3 of the complete alignment fully resolved the same topology as the MP rate analysis with even better support (BS support of ≥99 for all branches of the tree) (Figure 13). ML analysis of nucleotides from rate

category 4 of the gene alignment produced the same topology as previous ML analysis of this alignment, but phylogenetic resolution was not improved (Figure 14).

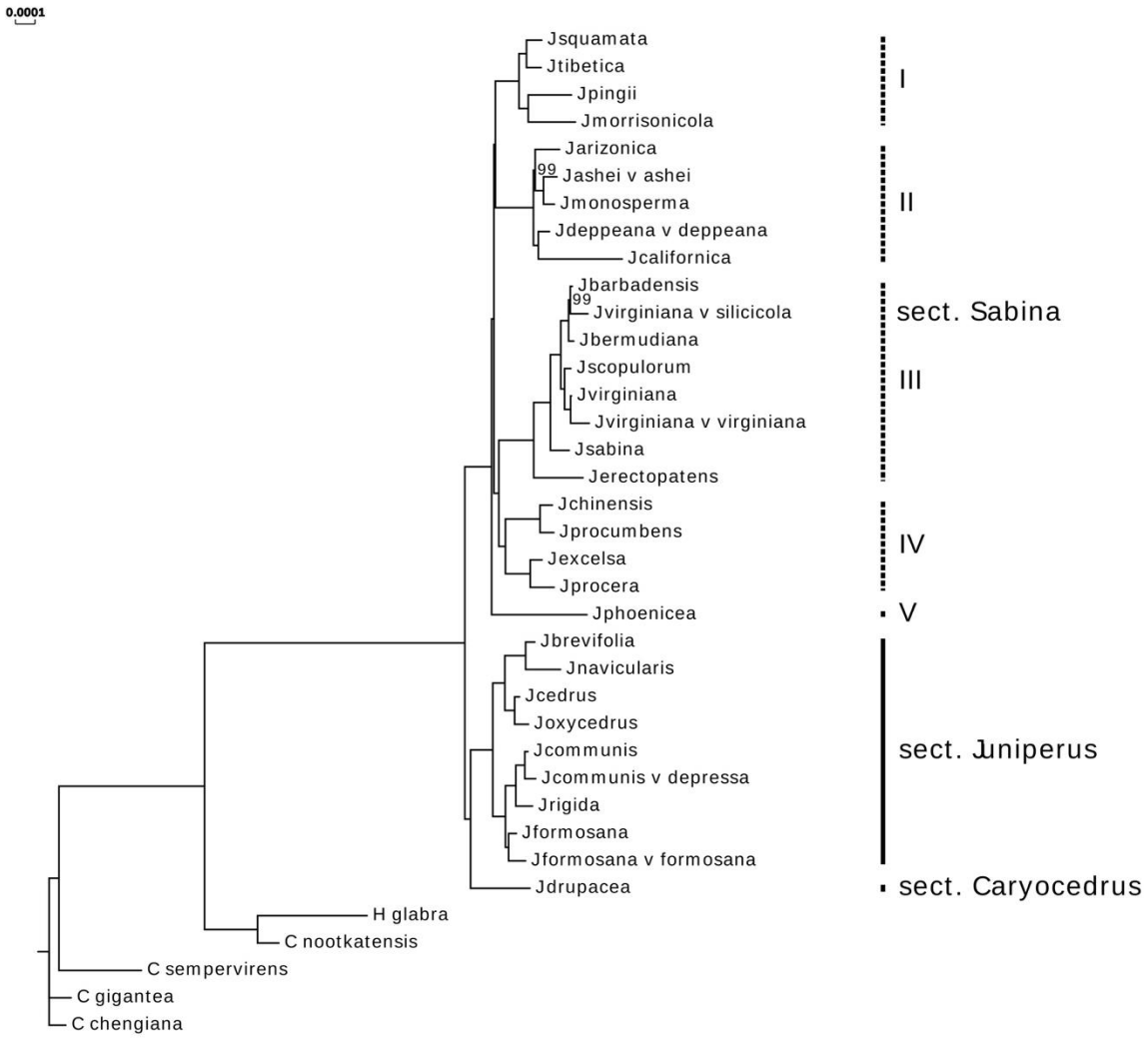


Figure 13. ML tree of complete-rate3 category. UfBS support values from 1000 replicates. All unlabeled nodes have 100 ultrafast bootstrap support.

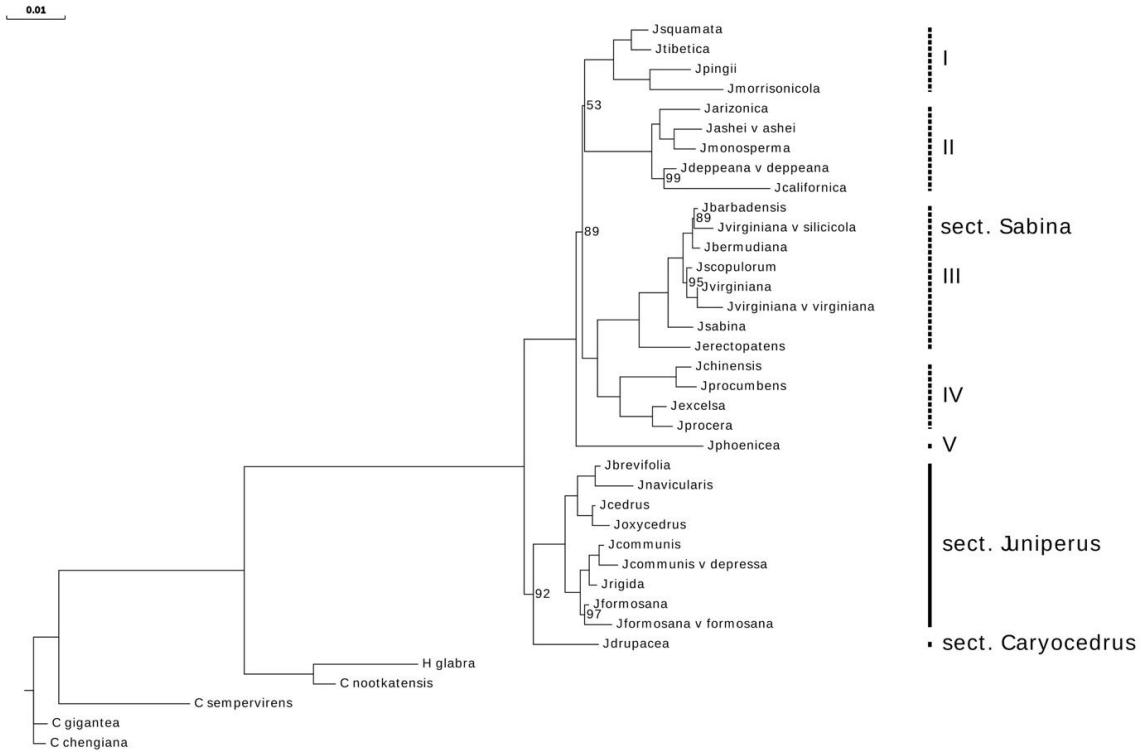


Figure 14. ML tree from gene-rate4 alignment. UfBS support values from 1000 replicates. All unlabeled nodes have 100 ultrafast bootstrap support.

*Hypothesis testing*

The AU test was used to evaluate the tree topologies generated during MP, ML, and BI analyses (Figure 3), as well as those proposed by Mao et al. (2010) and Adams and Schwarzbach (2013b), and a hybrid topology that has groups circumscribed as in Mao et al. (2010), but the relationships among groups presented in Adams and Schwarzbach (2013b). The ML tree topology, shared by the complete and gene-only analyses, was significantly better than all five alternative topologies (Figure 15), which were rejected with $p < 0.05$.

I also employed the AU test in an exploration of the sequence data used in Mao et al. (2010). The AU test was used to evaluate the ability of the data to reject alternative hypotheses in favor of a single best topology. The best tree, identified by the AU test, is

the same topology resolved by MP analysis of the gene-only alignment in our analysis (Figure 3C). While this topology had the lowest likelihood score, it was not significantly better than four of the remaining five topologies tested. Only one topology was rejected by the Mao et al. (2010) data, the topology proposed in Adams and Schwarzbach (2013b).

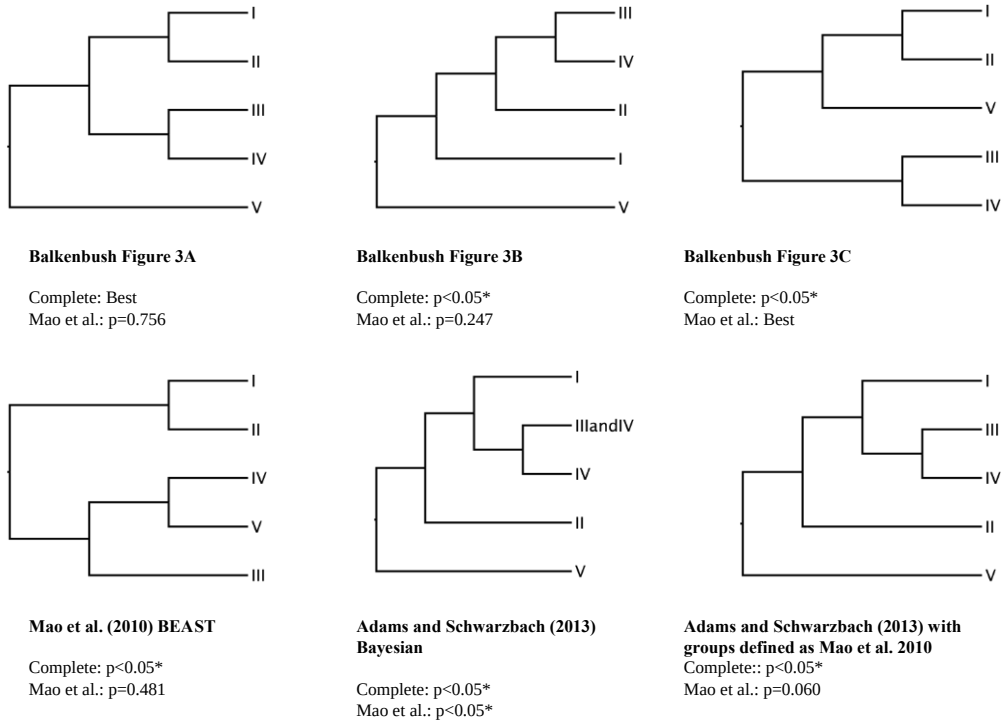


Figure 15. Trees used in hypothesis testing- AU test with complete matrix from the current study and data from Mao et al. (2010). * = topology rejected in favor of "best" scoring tree

# CHAPTER IV

# DISCUSSION

The two most recent phylogenies proposed for *Juniperus* are in conflict, specifically within the largest (60 species) and most diverse clade, sect. *Sabina* (Mao et al. 2010, Adams and Schwarzbach 2013b). The data used to generate these earlier phylogenies included 3,100 bp and 10,000 bp of chloroplast DNA, approximately 4-8% of the chloroplast genome. We present results of the first plastome-scale *Juniperus* phylogeny. The size of our data set, combined with a comprehensive phylogenetic analysis, provides strong resolution and support for an improved phylogenetic understanding of the backbone relationships of the genus.

Our topology agrees, in part, with each of the most recent topological hypotheses (Mao et al. 2010; Adams and Schwarzbach 2013b). The division of *Juniperus* into the three major sections: *Caryocedrus*, *Juniperus*, and *Sabina*, with sect. *Juniperus* and sect. *Caryocedrus* as sister clades, is consistent among all three studies, ours and the works of Mao et al. (2010) and Adams and Schwarzbach (2013b). The monophyly of the five groups in sect. *Sabina*, described by Mao et al. (2010), as well as the sister relationship between groups I and II in sect. *Sabina* are supported by our phylogenetic analyses, and results of the AU test. Although groups III and IV are not monophyletic in Adams and Schwarzbach (2013b), they do form a clade, and this grouping of species agrees with the sister relationship between groups III and IV resolved in our topology. The basal placement of group V in sect. *Sabina* is also in agreement with Adams and Schwarzbach (2013b).

The short branches estimated in our topology likely posed challenges in earlier attempts to fully resolve a chloroplast phylogeny of *Juniperus*. This genus experienced periods of both slow and rapid speciation. The stem lineage of *Juniperus* is well diverged from *Cupressus* indicating a long period lacking diversification, 15-20 My estimated by Mao et al. 2010, or the extinction of all but a single lineage. Likewise, the ancestral line of sect. *Sabina* experienced a relatively long period of divergence from the other two sections, *Caryocedrus* and *Juniperus*, before its subsequent diversification into the extant taxa (Figure 6). Unlike this slow, early evolution of the genus, sect. *Sabina* experienced initial branching events in rapid succession. Group V split quite early from the stem lineage of sect. *Sabina*, which then diversified rapidly into the two ancestral lineages of groups I and II and groups III and IV. The short internal branches of sect. *Sabina* (Figure 6) indicate that the early diversification events left little evidence in the DNA of the chloroplast genome, and recent mutations have likely obscured some of the older phylogenetic signature.

Mao et al. (2010) attempted the first comprehensive *Juniperus* chloroplast phylogeny, with 77 juniper accessions, representing 51 species, and 39 outgroup accessions. The 10,299-character data matrix, roughly 8% of the cp genome, contained 1,173 parsimony informative characters (PICs), including indels. Despite this relatively large amount of sequence data, it proved insufficient to resolve the rapid diversification of early lineages within sect. *Sabina*.

The use of complete chloroplast data, with adequate taxon sampling, is likely the reason for our ability to fully resolve the relationships within *Juniperus*, a large genus with a complex evolutionary history. Other researchers of genera containing rapid

radiations have identified the necessity for complete chloroplast genomes to capture adequate variation among species, but at times, even complete genomes lack sufficient variation to provide complete resolution (Whittall et al. 2010; Parks et al. 2009). Slightly more than one third (28 of 75) of all recognized *Juniperus* species were included in our analyses. All major groups of *Juniperus* are represented, with several taxa representing larger clades, and taxa sampled from regions of conflict between the two recently published phylogenies. In an evaluation of numerous genus-level chloroplast-based phylogenies, Parks et al. (2009) identified a significant relationship between increased resolution and increase in matrix length, but no correlation between improved resolution with increased taxa sampling. It is unlikely, given our sampling coverage, that inclusion of additional taxa would resolve more accurate backbone relationships in the chloroplast phylogeny of *Juniperus*.

Comparison of our results with those of Mao et al. (2010) and Adams and Schwarzbach (2013b) revealed that the placement of group V within sect. *Sabina* has been problematic. While group V was primarily resolved as the basal clade sect. *Sabina* during our analyses, and the analysis of Adams and Schwarzbach (2013b), it has alternate placements in the BEAST analysis (implemented in a Bayesian framework) of Mao et al. (2010), and in the Bayesian analysis of our complete, un-partitioned data set. The placement of clade V poses a challenge because it contains only one extant taxon, *J. phoenicea*, and it is well diverged from the other clades in the section. Inaccurate relationships can be resolved because of the accumulation of mutations over a long divergence time, which leads to long-branch attraction. This could occur if mutations in

*J. phoenicea* appear shared with other clades, but are homoplasious, and only a matter of coincidence.

While the alternate placements of group V may be due to long-branch-attraction, it is interesting to note, that these placements occurred within the Bayesian analyses. Mao et al. (2010) do not report support values for their BEAST analysis, so we don't know how well-supported their alternative placement of group V was. The placement of group V, sister to groups (I,II), produced by our Bayesian analysis was well supported by posterior probabilities (Figure 9), but posterior probabilities have been found to "substantially" over-estimate support (Suzuki et al. 2002; Erixon et al. 2003; Stull et al. 2015), while ML bootstrap support values are considered more conservative. The AU test, implemented in PAUP, under likelihood settings, rejected the alternative placement of group V, in favor of our predominant topology with the basal placement of group V. We are confident in this basal placement of group V, which was resolved in 22 of our 25 analyses, and underline the importance of utilizing a variety of phylogenetic methods and evaluating the strength of the signal in the data by its ability to reject alternative topological hypotheses.

We observed alternative topologies produced by Bayesian analysis of the chloroplast genome from different subsets of data and partitioning schemes. Adams and Schwarzbach (2013b) produced a well-supported topology by Bayesian Inference analysis of cpDNA and nrITS. The topology among groups within sect. *Sabina* differs from both ours and Mao et al. (2010). It was not possible to explore the data using other phylogenetic methods, as done with the data of Mao et al. (2010), because the data was not publicly available at the time of this study. Their topology is well-supported by

posterior probabilities, which are not directly comparable with nonparametric bootstrap support values from ML analysis, as discussed above. Therefore, we do not know how the topology would be supported in a likelihood framework, nor the ability of the data set to reject alternative hypotheses. The AU tests of our complete alignment, as well as the Mao et al. (2010) data set both rejected this topology in favor of higher scoring topologies.

Degree of resolution and bootstrap support can be affected by the presence of signal noise in phylogenetic data sets. Strategies used to remove data most prone to signal noise or substitution saturation, include removing third codon nucleotides, which often have higher mutation rates, or removing more quickly evolving genes themselves. These methods can increase the signal to noise ratio in a data set and improve the measured support of phylogenies (Philippe et al. 2000; Nozaki et al. 2007; Klopfstein et al. 2017). A side effect of these practices is the loss of phylogenetic information. All third codon positions do not exhibit substitution saturation and many positions in a quickly evolving gene will contain valuable phylogenetic information. A strategy to reduce the loss of signal, when eliminating noise, is to filter nucleotide positions by their evolutionary rates, eliminating both rapidly evolving, and constant or near constant sites. IQTREE calculates evolutionary rates of each nucleotide position, and sorts them into the number of rate categories empirically determined during the model-fitting process.

We performed MP and ML analyses on the most phylogenetically informative nucleotide rate category from each of our alignments. The use of this strategy, dramatically improved the BS support in both MP and ML analyses of the complete alignment, producing a tree with all but one node with ≥ 95 BS support in the MP

analysis and all nodes ≥ 95 BS support in the ML analysis. The analysis of this subset of the complete alignment improved the main topology of our phylogenetic analyses, which was already relatively well supported.

A factor affecting our ability to compare our phylogeny to that of Adams and Schwarzbach (2013b) is the use of different DNA sources. The Adams and Schwarzbach phylogeny presents results of combined nuclear-ribosomal ITS (nrITS) DNA and cpDNA. Mao et al. (2010) identified significant conflict in phylogenetic signal between nrITS and cpDNA in *Juniperus*. Conflict in DNA sourced from nuclear and organellar genomes can result from their different evolutionary histories. With hybridization known to occur in the genus, the uniparentally-inherited chloroplast may support a different phylogeny than the bi-parentally inherited nuclear genome. Risks have been associated with the use of nrITS for phylogenetic inference (Alvarez and Wendel 2003), and single-copy-nuclear (SCN) genes are advocated as a less homoplasious source for nuclear phylogenetic markers. Further research by Adams (2015) and Adams et al. (2016, 2017) found evidence of concerted evolution in nrITS, indicated by hybrid individuals with copies of the SCN gene maldehey from each parent, but nrITS copies from only one parent. Further phylogenetic investigation of *Juniperus* will need to employ several SCN genes, as it is unlikely that only one or two will accurately trace the evolution within the genus.

In addition to further phylogenetic exploration, additional investigation into the biogeographic history of *Juniperus* could be pursued. Mao et al. (2010) executed several biogeographic and molecular dating analyses utilizing a sample of the chloroplast genome. Future work could expand upon the biogeographic and molecular dating

analyses conducted in Mao et al. (2010), possibly under the assumption of our plastome-scale phylogeny, or incorporating evidence from several SCN genes. The genus has a very wide geographic distribution, and questions concerning its colonization patterns and divergence dates remain to be answered.

## REFERENCES


Adams RP. 1994. Geographic variation in the volatile terpenoids of *Juniperus monosperma* and *J. osteosperma*. Biochemical Systematics and Ecology 22:65-72.

Adams RP. 2004. Junipers of the World: the genus *Juniperus*. Vancouver, BC, Canada: Trafford Publishing.

Adams RP. 2008. Junipers of the World: the genus *Juniperus*, 2nd edn. Vancouver, BC Canada: Trafford Publishing.

Adams RP. 2014. Junipers of the World: the genus *Juniperus*, 4th edn. Vancouver, BC Canada: Trafford Publishing.

Adams RP. 2015. Allopatric hybridization and introgression between *Juniperus maritima* R. P. Adams and *J. scopulorum* Sarg. II. Additional Evidence from nuclear and cpDNA genes in Montana, Wyoming, Idaho and Utah. Phytologia 97(3):189-199.

Adams RP, Demeke T. 1993. Systematic relationships in *Juniperus* based on random amplified polymorphic DNAs (RAPDs). Taxon 42:553-572.

Adams RP, Nguyen S, Liu J. 2006. Geographic variation in the leaf essential oils of *Juniperus sabina* and var. *arenaria*. Journal of Essential Oil Research. 18:497-502.

Adams RP, Schwarzbach AE. 2012a. Taxonomy of *Juniperus* section *Juniperus*: Sequence analysis of nrDNA and five cpDNA regions. Phytologia 94(2):280-297.

Adams RP, Schwarzbach AE. 2012b. Taxonomy of the multi-seeded, entire leaf taxa of *Juniperus* section *Sabina*: Sequence analysis of nrDNA and four cpDNA regions. Phytologia 94(3):350-368.

Adams RP, Schwarzbach AE. 2012c. Taxonomy of the turbinate seed cone taxa of *Juniperus* section *Sabina*: Sequence analysis of nrDNA and four cpDNA regions. Phytologia 94(3):388-403.

Adams RP, Schwarzbach AE. 2013a. Taxonomy of the serrate leaf *Juniperus* of North America: Phylogenetic analyses using nrDNA and four cpDNA regions. Phytologia 95(2):172-178.

Adams RP, Schwarzbach AE. 2013b. Phylogeny of *Juniperus* using nrDNA and four cpDNA regions. Phytochemistry 95:179-187.

Adams RP, Armagan M, Boratynski A, Douaihy B, Dagher-Kharrat MD, Farzaliyev V, Gucel S, Mataraci T, Tashev AN, Schwarzbach AE. 2016. Evidence of relictual introgression or incomplete lineage sorting in nrDNA of *Juniperus excelsa* and *J. polycarpos* in Asia Minor. Phytologia 98(2):146-155.

Adams RP, Gonzalez-Elizondo MS, Gonzalez-Elizondo M, Noy DR, Schwarzbach AE. 2017. DNA sequencing and taxonomy of unusual serrate *Juniperus* from Mexico: Chloroplast capture and incomplete lineage sorting in *J. coahuilensis* and allied taxa. Phytologia 99:62-73.

Alvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution. 29(3):417-34.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 30:2114–2120.

Boratyński A, Wachowiak W, Dering M, Boratyńska K, Sekiewicz K, Sobierajska K, Jasińska AK, Klimko M, Maria-Montserrat J, Romo A, Ok T, Didukh Y. 2014. The biogeography and genetic relationships of *Juniperus oxycedrus* and related taxa from the Mediterranean and Macaronesian regions. Botanical Journal of the Linnean Society 174(4):637-653.

Bushnell B. Unpublished. BBMap. sourceforge.net/projects/bbmap/

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Systematic Biology. 65:997-1008.

Cronn R, Liston A, Parks M, Gernadt DS, Shen R, Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Research. 36(19):e122

Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K. 2011. The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Research 18(2):93–105.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15.

Doyle JJ. 1992. Gene trees and species trees: molecular systematics as one- character taxonomy. Systematic Botany 17:144–163.

Eckenwalder, JE. 2009. Conifers of the world: the complete reference. Portland: Timber Press.

Endlicher SL. 1847. Synopsis Coniferarum. Scheitlin und Zollifer, Sangalli (Sankt Gallen).

Erixon P, Svennbald D, Britton T, Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Systematic Biology. 52(5):665-673.

Farjon A. 2005. A monograph of Cupressaceae and Sciadopitys. Kew, UK: Royal Botanic Gardens, Kew.

Fassett NC. 1944a. *Juniperus virginiana, J. horizontalis* and *J. scopulorum*-I. The specific characters. Bulletin of the Torrey Botanical Club. 71(4):410-418.

Fassett NC. 1944b. *Juniperus virginiana, J. horizontalis* and *J. scopulorum* -II. Hybrid swarms of *J. virginiana* and *J. horizontalis*. Bulletin of the Torrey Botanical Club. 71:475-483.

Fasset NC. 1945a. *Juniperus virginiana, J. horizontalis* and *J. scopulorum* -III. Possible hybridization of *J. horizontalis* and *J. virginiana*. Bulletin of the Torrey Botanical Club. 72:42-46.

Fasset NC. 1945b. *Juniperus virginiana, J. horizontalis* and *J. scopulorum* -IV. Hybrid swarms of *J. virginiana* and *J. horizontalis*. Bulletin of the Torrey Botanical Club. 72:379-384.

Fassett NC. 1945c. *Juniperus virginiana, J. horizontalis* and *J. scopulorum*. V. Taxonomic treatment. Bulletin of the Torrey Botanical Club 72:480-482.

García D. 2001. Effects of seed dispersal on *Juniperus communis* recruitment on a Mediterranean mountain. Journal of Vegetation Science. 12:839–848.

Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP. 2014. Predominant and substoichiometric isomers of the plastid genome

coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. Genome Biology and Evolution 6(3):580-590.

Hall MT. 1952. A hybrid swarm in Juniperus. Evolution 6(4):347-366.

Hughes CE, Eastwood RJ, Bailey CD. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 361:211–25.

Jordano P. 1993. Geographical ecology and variation of plant-seed disperser interactions: Southern Spanish junipers and frugivorous thrushes. Vegetatio, 107/108:85–104.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods. 14:587-589.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution. 30(4):772–780.

Klopfstein S, Massingham T, Goldman N. 2017. More on the best evolutionary rate for phylogenetic analysis. Systematic Biology. 66(5):769-785.

Krussmann G. 1991. Manual of cultivated conifers. Timber Press. Portland, OR.

Lawrence BM. 1985. A review of the world production of essential oils (1984). Perfumer and Flavorist 10:1-16.

Lemoine-Sebastian C. 1967. Appareil reproducteur male des Juniperus. Trav. Lab. Forest. Toulouse, T.1, vol. 6, art. 29:1-35.

Lemoine-Sebastian, C. 1968. L'inflorescence femelle des Juniperae: ontogenese, structure, phylogenese. Trav. Lab. Forest. Toulouse, Tom. 1, vol. 7, art. 5: 1-456.

Linnaeus C. 1753. Species Plantarum. Laurentius Salvius, Stockholm.

Little DP. 2006. Evolution and circumscription of the true cypresses (Cupressaceae: *Cupressus*). Systematic Botany 31:461–480.

Liu J, Chen Z, Lu A. 2002. Molecular evidence for the sister relationship of the eastern Asia-North American intercontinental species pair in the *Podophyllum* group (Berberidaceae). Botanical Bulletin of Academia Sinica 43:147-154.

Livingston RB. 1972. Influence of birds, stones and soil on the establishment of pasture juniper, *Juniperus communis*, and red cedar, *J. virginiana* in New England Pastures. Ecology 53(6):1141–47.

Mao K, Hao G, Liu J, Adams RP, Milne RI. 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): Variable diversification rates and multiple intercontinental dispersals. New Phytologist 188(1):254-272.

Mao K, Milne RI, Zhang L, Peng Y, Liu J, Thomas P, Mill RR, Renner SS. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proceedings of the National Academy of Sciences 109(20):7793-7798.

Mihn BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution. 30(5):1188-1195.

Morrel JJ. 2011. Resistance of selected wood-based materials to fungal and termite attack in non-soil contact exposures. Forest Products Journal 61(8):685-687.

Morrell JJ, Miller DJ, Schneider PF. 1999. Service life of treated and untreated fence posts: 1996 post-farm report. Corvallis, Oregon. College of Forestry, Forest Research Laboratory, Oregon State University.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ 2015 IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Molecular Biology and Evolution 32:268-274.

Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N, Watanabe M. 2007. Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. Molecular Biology and Evolution 24:1592–1595.

Parks M, Cronn R, Liston A. 2009 Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BioMed Central. 7:84.

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Müller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proceedings of the Royal Society of London. B Biological Sciences. 267:1213–1221.

Raubeson LA, Jansen RK. 2005. Chloroplast genomes of plants. In: Henry RJ, editor. Plant diversity and evolution: Genotypic and phenotypic variation in higher plants. Cambridge, MA: CABI. p. 45–68.

Rieseberg LH, Soltis DE, 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants. 5:65–84.

Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA. Huelsen- beck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539-542.

Rumeu B, Caujape-Castells J, Blanco-Pastor JL, Jaen-Molina R, Nogales M, Elias RB, Vargas P. 2011. The colonization history of *Juniperus brevifolia* (Cupressaceae) in the Azores Islands. PloS one 6(11):1-13.

Santos T, Telleria JL, Virgos E. 1999. Dispersal of Spanish juniper *Juniperus thurifera* by birds and mammals in a fragmented landscape. Ecography 2:193-204.

Schupp EW, Gómez JM, Jiménez JE, Fuentes M. 1997. Dispersal of *Juniperus occidentalis* (western juniper) seeds by frugivorous mammals on Juniper Mountain, southeastern Oregon. Great Basin Naturalist 57:74-78.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology. 51:492-508.

Spach, E. 1841. Révision des Juniperus. Annales des sciences naturelles. Botanique. 2(16):282-305.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. Systematic Biology. 75(5):758-771.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn R, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany 99:349 – 364.

Stull GW, Duno de Stefano R, Soltis DE, Soltis PS. 2015. Resolving the basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. American Journal of Botany. 102(11):1794-1813.

Suzuki Y, Glazko GV. Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proceedings of the National Academy of Sciences U.S.A. 99:16138–16143.

Swofford, DL. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Teixeira H, Rodríguez-Echeverría S, Nabais C. 2014. Genetic diversity and differentiation of *Juniperus* thurifera in Spain and Morocco as determined by SSR. PloS One 9(2):e88996.

Thorne RF. 1972. Major disjunctions in the geographic ranges of seed plants. The Quarterly Review of Biology 47(4):365–411.

Whittall JB, Syring J, Parks M, Benrostro J, Dick C, Liston A, Cronn R. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Molecular Ecology. 19:100-114.

Wu C-S, Chaw S-M. 2014. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): Evolution towards shorter intergenic spacers. Plant Biotechnology Journal. 12(3):344–353.

Yang ZY, Ran JH, Wang XQ. 2012. Three genome-based phylogeny of Cupressaceae s.l.: further evidence for the evolution of gymnosperms and Southern Hemisphere biogeography. Molecular Phylogenetics and Evolution. 64:452-470.

Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. 2015. The first complete chloroplast genome sequences in Actinidiaceae: Genome structure and comparative analysis." Plos One 10(6):e0129347.

Zamora R. 1990. The fruit diet of ring ouzels *Turdus torquatus* wintering in the Sierra Nevada southeast Spain. Alauda. 58(1):67-70.