Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution xxx (2008) xxx–xxx
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

A molecular phylogeny of the wild onions (*Allium*; Alliaceae) with a focus on the western North American center of diversity

Nhu H. Nguyen, Heather E. Driscoll, Chelsea D. Specht *

Department of Plant and Microbial Biology, University of California, Berkeley, 111 Koshland Hall, MC 3102, Berkeley, CA 94720, USA

Received 20 September 2007; revised 4 December 2007; accepted 4 December 2007

Abstract

Nuclear ribosomal DNA (ITS and ETS) sequences from 39 native Californian (USA) *Allium* species and congeners were combined with 154 ITS sequences available on GenBank to develop a global *Allium* phylogeny with the simultaneous goals of investigating the evolutionary history (monophyly) of *Allium* in the Californian center of diversity and exploring patterns of adaptation to serpentine soils. Phylogenies constructed with ITS alone or ITS in combination with ETS provided sufficient resolution for investigating evolutionary relationships among species. The ITS region alone was sufficient to resolve the deeper relationships in North American species. Addition of a second marker (ETS) further supports the phylogenetic placements of the North American species and adds resolution within subgenus *Amerallium*, a clade containing many Californian endemics. Within the global phylogeny, the native North American species were found to be monophyletic, with the exception of *Allium tricoccum* and *Allium schoenoprasum*. All native Californian species included in the analysis fell into a monophyletic subgenus *Amerallium* section *Lophioprason*, although endemic Californian species were not monophyletic due to the inclusion of species with ranges extending beyond the California Floristic Province. The molecular phylogeny strongly supports previous morphology-based taxonomic groupings. Based on our results, serpentine adaptation appears to have occurred multiple times within section *Lophioprason*, while the ancestor of the Californian center of diversity may not have been serpentine-adapted.

Keywords: Wild onion; *Allium*; ETS; ITS; Molecular phylogeny; Ecological reconstruction; *Amerallium*; *Lophioprason*; Serpentine; Californian endemic

1. Introduction

The California Floristic Province (CFP) is a Mediterranean-type ecosystem in western North America that is extremely varied in topography, microclimate, geology and soils resulting in a high percentage of endemism of its plants and animals. As one of only five Mediterranean climate areas in the world, the CFP has been classified as a Biodiversity Hotspot (Conservation International, 2007) and is the center of diversity in North America for many groups of plants, including the genus *Allium* L.

Allium comprises a major taxonomic portion of the monocot family Alliaceae and consists of a diverse group of perennial herbs characterized by rhizomatous or, more commonly, bulbous stems (Fig. 1; M, N), narrow basal leaves, flowers with six free tepals (Fig. 1; A, C, G–I) and superior ovaries with 0–6 crests (Fig. 1; C, H), inflorescences of scapose umbels (Fig. 1; A–L), mucilaginous latex, and a distinctive onion-like odor and taste due to the presence of cystine sulfoxides. *Allium* contains many economically important crop and ornamental species, including onion (*Allium cepa*), garlic (*A. sativum*), leek (*A. ampeloprasum*), and chive (*A. schoenoprasum*). *Allium* is naturally distributed throughout the Northern Hemisphere and is represented by ca. 750 species worldwide (Stern, 1992) with ca. 100 species in North America (McNeal and Jacobsen, 2002) and ca. 50 species and 20 varieties native to California (McNeal, 1992; Hickman, 1993). The genus is mainly restricted to regions that are seasonally dry, with centers of diversity in southwest/central Asia, eastern Asia, and

* Corresponding author. Fax: +1 510 642 4995.
E-mail address: cdspecht@nature.berkeley.edu (C.D. Specht).

1055-7903/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2007.12.006

Fig. 1. Selected images of *Allium* (section *Lophioprason*) species used in this study. (A) *A. falcifolium* inflorescence and falcate leaves, (B) *A. lemmonii*, a species in subsection *Falcifolia*, (C) *A. campanulatum* showing the dark ovary crests in the center, (D) *A. sanbornii*, a representative of subsection *Sanborniana*, (E) *A. sharsmithiae*, a serpentine endemic species showing the umbel inflorescence, (F) *A. diabolense*, a species in subsection *Sanborniana*, (G) *A. shevockii*, a species in subsection *Sanborniana*, (H) *A. haematochiton* (serpentine form), (I) *A. haematochiton* (non-serpentine form) highlighting the pink ovary crests in the center, (J) *A. amplectens* (non-serpentine form), (K) *A. unifolium*, a representative in subsection *Acuminata*, (L) *A. bolanderi*, a species in subsection *Bolanderiana*, (M) *A. bolanderi* bulbs and rhizomes, (N) *A. serra* ("herringbone" pattern in bulb scale), (O) *A. fimbriatum* var. *purdyi* growing on serpentine soil.

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

in North America. The two North American centers of diversity occur in Texas (ca. 15 species + 5 varieties) and the California Floristic Province.

Characters including nectary morphology, leaf number and shape/size, ovary crest and seed morphology, inflorescence structure, vegetative anatomy, basic chromosome number, and bulb/rhizome morphology in combination with biogeographic patterns have been used to place species into subgenera and sections. Many of these groupings have been supported and refined with recent studies using chloroplast and nuclear (ITS) genomic data (Samoylov et al., 1995; Linne von Berg et al., 1996; Dubouzet and Shinoda, 1998, 1999; Mes et al., 1999; Samoylov et al., 1999; Friesen et al., 2006).

The taxonomy of *Allium* is complicated with a great number of synonyms and intrageneric groupings (reviewed in Klaas, 1998). A recent phylogenetic analysis of *Allium* using ITS sequence data and representing one-quarter of the diversity in the genus tested the monophyly of *Allium* and evaluated the relationships among intrageneric groups (subgenera and sections) previously circumscribed (Friesen et al., 2006). Their analysis provided support for a monophyletic *Allium* s.l. (including *Milula* and *Nectaroscordum*) but found several of the traditional subgenera to be non-monophyletic. The authors proposed a new intrageneric classification with divisions into subgenera and sections based on their ITS phylogeny.

Although Friesen et al. (2006) include species from all major taxonomic groups in the genus in their phylogeny, most *Allium* species sampled are from the Old World, with only six New World species represented: *A. amplectens* Torrey native to the Pacific coast of North America (section *Lophioprason*), *A. cernuum* Roth native to western North America (section *Lophioprason*), *A. unifolium* Kellogg native to California and Oregon (section *Lophioprason*), *A. brevistylum* S. Watson from Utah (section *Caulorhizideum*), *A. goodingii* Ownbey from Arizona (section *Caulorhizideum*), and *A. glandulosum* Link & Otto native to the southwestern United States (section *Rhopetoprason*). *Allium fimbriatum* S. Watson is listed in the taxon sampling but is not included in the phylogenetic analysis (Friesen et al., 2006; their Fig. 1). This species traditionally has been placed in section *Rhopetoprason* along with *A. glandulosum*. In their analysis, the New World species fell into a single clade corresponding to subgenus *Amerallium*. *Amerallium* contains both New World and Old World species and was redefined in Friesen et al. (2006) to exclude sections *Microscordum* and *Nectaroscordum*, with the former placed sister to *Amerallium* and the latter placed sister to the *Microscrodium* + *Amerallium* clade. In their analysis, section *Microscordum* was represented by *Allium monanthum* while section *Nectaroscordum* was represented by *A. bulgaricum*. Within *Amerallium*, North American species were placed in three closely related sections: *Lophioprason* (3 spp.), *Rhopetoprason* (1 sp.), and *Caulorhizideum* (1 sp.). When tested, these sections were found to be monophyletic and they grouped together to form a well-supported North American clade within *Amerallium* (Friesen et al., 2006).

Subgenus *Amerallium* was described by Traub (1968) to include three isolated geographical groups, one containing almost all *Allium* species native to North America (New World) and a second containing a smaller number of Mediterranean/North African and eastern Asian (Old World) species. Both groups were found to have a basic chromosome number of $n = 7$ and have been shown in subsequent studies to indeed represent two distinct biogeographic clades (New World and Old World) within a monophyletic *Amerallium* (Samoylov et al., 1995; Dubouzet and Shinoda, 1999). Dubouzet and Shinoda (1999) included 10 New World species from sections *Lophioprason* and *Caulorhizideum* in their analysis of relationships among Old and New World taxa focusing on Traub's *Amerallium* (1968, 1972). Sections *Lophioprason* and *Caulorhizideum* were reciprocally monophyletic; however, *Lophioprason* alone contains a large diversity of species divided by Ownbey into five alliances (*A. acuminatum, A. campanulatum, A. cernuum, A. falcifolium*, and *A. sanbornii*) or Traub into subsections (*Acuminata, Bolanderiana, Campanulata, Cernua, Falcifolia*, and *Sanborniana*). Dubouzet and Shinoda (1999) suggested the use of genetic distances for possible reclassification and designation of rank within *Amerallium*, and suggested the elevation of some subsections (*Acuminata*, *Cernua*, *Sanborniana*, and possibly *Bolanderiana*) to sections. The reported genetic distances demonstrated a high degree of molecular diversity among North American alliums, however the sampling was limited for this group (10 species total). In order to deal with this taxonomic and genetic diversity, they recommended dividing section *Lophioprason* into six independent sections based on Ownbey's alliances (Saghir et al., 1966; see McNeal, 1992) or elevating Traub's subsections (Traub, 1968), and separating Old and New World *Amerallium* into two different subgenera, *Molium* and *Amerallium*, respectively.

The primary goal of the research presented here is to expand on the existing *Allium* molecular phylogeny by including species from western North America with a focus on *Amerallium* section *Lophioprason*, including multiple representatives from each of Traub's subsections (as corresponding to Ownbey's alliances). We use nuclear ribosomal DNA sequence data from the internal transcribed spacer (ITS) and external transcribed spacer (ETS) from Californian *Allium* species and congeners to develop a global *Allium* phylogeny. Furthermore, we use this phylogeny to investigate the evolutionary history (monophyly) of alliums in western North America, especially in California, and begin to investigate patterns of adaptation to serpentine soil (ultramafic soils which have a skewed ratio of Ca:Mg toxic to many plants), an important edaphic factor in studies of plant adaptation, speciation, and diversification across the California Floristic Province (Clausen et al., 1948; Kruckeberg, 1984; O'Dell and Claassen, 2006; Wright et al., 2006).

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

2. Material and methods

2.1. Taxon sampling

A total of 223 taxa (excluding duplicates) were included in the study group for the global single-marker phylogenetic analysis. The ingroup comprised 212 *Allium* species (213 taxa including varieties) with both New and Old World distributions, including multiple members of many of the published *Allium* sections (see Klaas, 1998; Friesen et al., 2006). *Tulbaghia fragrans*, *Ipheion uniflorum*, five *Dichelostemma* species, and two *Nothoscordum* species were designated as outgroups. These nine outgroup species were chosen using cladistic groupings in previous phylogenetic analyses (Friesen et al., 2006). Seventy-one taxa (69 species plus two varieties) were included in the study group for the combined nrDNA (ITS and ETS) analyses, with 62 ingroup taxa and the same nine outgroup species as listed above. A complete list of taxa used in the study, with GenBank accession numbers for each marker and voucher information, is provided in Table 1. Material for DNA extraction was of wild origin, collected from botanical gardens (UCBG, RPBG, and Rancho Santa Ana), or from herbarium sheets from the University and Jepson Herbaria at the University of California, Berkeley (UC and JEPS, respectively) (Table 1). The freshly collected material was preserved in silica gel and stored at −80 °C until extraction. Vouchers of wild-collected specimens were deposited into UC and JEPS. Sequences generated for this analysis are deposited in GenBank and can be found in Table 1; sequences previously accessioned to GenBank and included in the analysis are listed at the end of the table.

2.2. DNA extraction, PCR amplification, and sequencing

Total genomic DNA was extracted from leaf samples using the commonly used CTAB method (Doyle and Doyle, 1987). Both nrDNA markers, ITS and ETS, were amplified using the polymerase chain reaction (PCR) in 25 μL aliquots with the following reaction components: 50–150 ng of genomic DNA, 1X iProof buffer (Bio-Rad, Hercules, CA, USA), 200 μmol/L of each dNTP, 0.1 μmol/L of each primer, and 0.625 U iProof polymerase (Bio-Rad) using a Bio-Rad thermal cycler (MyCycler) with thermal cycling conditions as follows: for ITS, initial denaturation was for 3 min at 98 °C, followed by 35 cycles of 15 s at 98 °C, 30 s at 59 °C, 30 s at 72 °C, finishing with a final extension of 72 °C for 7 min; for ETS initial denaturation was for 5 min at 95 °C, followed by 40 cycles of 1 min at 94 °C, 1 min at 50 °C, 2 min at 72 °C, finishing with 72 °C for 8 min. ITS segments were amplified using primers ITS Leu (Baum et al., 1998) and ITS4 (White et al., 1990); and *Allium*-specific primers ITS-A and ITS-B (Blattner et al., 2001). ETS segments were amplified using a forward primer designed for this study, *Allium* ETSF (TTC GTG GTY GGC GGA GCT), and 18S-IGS (Baldwin and Markos, 1998). Prior to sequencing, PCR products were purified with the enzymes Exonuclease I from *E. coli* and Shrimp Alkaline Phosphatase from *Padalis borealis* following the manufacturer's recommendations (Fermentas, Hanover, MD, USA). The purified PCR products were sequenced in both directions using the PCR primers and, in some cases, ITS-C and ITS-D and sequencing specific primers ITS-SF and ITS-SR (Blattner et al., 2001). The ABI Prism BigDye Terminator Cycle Sequence Ready Reaction Kit v.3.1 (Perkin-Elmer/Applied Biosystems, Foster City, CA, USA) sequencing chemistry was used with the following thermal cycling parameters: initial denaturation was for 5 min at 80 °C, followed by 30 cycles of 10 s at 96 °C, 5 s at 50 °C, and finishing with 50 °C for 4 min. Sequencing products were resolved on an ABI Prism 3100 automated sequencer (Applied Biosystems, Foster City, CA). Because ETS was difficult to directly sequence for *Allium dichlamydeum*, we cloned ETS for this species. Unpurified ETS templates were ligated into the vector from the Gene JET™ PCR Cloning Kit following the manufacturer's recommendations (Fermentas, Hanover, MD, USA) and subsequently transformed into chemically competent *E. coli*. After overnight culture at 37 °C on LB–ampicillin selective plates, colonies carrying the ETS insert were arbitrarily selected and used as a template for PCR amplification using the protocol outlined above.

2.3. Sequence alignment and indel coding

Raw forward and reverse sequences for each sample were assembled, ambiguous bases were corrected by inspection of chromatograms, and consensus sequences were edited using Sequencher (version 4.7; Gene Codes Corporation, Ann Arbor, MI, USA). Consensus sequences for each region were aligned manually using MacClade v.4.06 (Maddison and Maddison, 2003) and PAUP* version 4.0b10 (Swofford, 2003). Where insertion–deletions (indels) were shared by two or more taxa and could be aligned unequivocally, they were treated as potentially phylogenetically informative. Indels were coded as independent, single, binary characters following Simmons and Ochoterena (2000). Once indels were coded, all gap regions were treated as missing data. Corrected pairwise sequence divergence per site for each nrDNA marker was calculated for the dataset based on the best-fit model of nucleotide sequence substitution as determined by a hierarchical likelihood ratio test in Modeltest v. 3.06 (Posada and Crandall, 1998). Aligned data for all analyses are available from the first author upon request.

2.4. Phylogenetic analyses

Phylogenetic analyses for the ITS-only and combined datasets were conducted employing maximum parsimony (MP) and Bayesian methods. For parsimony analyses, heuristic searches were performed on 1000 replicates with random taxon addition, 1 tree held during tree–bisection–reconnection (TBR) branch swapping for each addition

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

Table 1
Collection information for *Allium* species included in analyses

NN Col. #	Species and authority	Herbarium voucher no.	Living collection no.	GenBank ITS	GenBank ETS	Coll. locality
NN 115	*A. abramsii* (Ownbey & Aase ex Traub) McNeal	JEPS 88734	—	EU096131		Jepson Herbarium
NN 54	*A. amplectens* Torr.	—	UCBG 53.0624	EU096132		UC Botanical Garden
NN 85	*A. amplectens* Torr.	NN 85	—	EU096133	EU162704	UC McLaughlin Reserve, CA, USA; 38°51′52.23″N, 122°25′35.19″W
NN 93	*Allium anceps* Kellogg	JEPS 94130	—	EU096134	EU162705	Jepson Herbarium
NN 112	*Allium anceps* Kellogg var. *anceps*	UC 1732188	—	EU096135	EU162706	UC Herbarium
CS 07-03	*Allium angulosum* L.	—	OZ 1995BL00214	EU096136		Utrecht BG
NN 92	*Allium atrorubens* S. Watson var. *atrorubens*	UC 173122	—	EU096137	EU162707	UC Herbarium
NN 101	*Allium atrorubens* S. Watson var. *cristatum* (S. Watson) McNeal	UC 96559	—	EU096138	EU162708	UC Herbarium
NN 72	*Allium bolanderi* S. Watson	—	RPBG 93.149	EU096139	EU162709	Regional Parks Botanical Garden
NN 88	*Allium bolanderi* S. Watson	NN 88	—	EU096140		UC McLaughlin Reserve, CA, USA; 38°51′23.77″N, 122°22′15.51″W
NN 56	*Allium bolanderi* S. Watson var. *bolanderi*	—	UCBG 83.0924	EU096141	EU162710	UC Botanical Garden
NN 114	*Allium burlewii* Davidson	UC 1584595	—	EU096142	EU162711	UC Herbarium
NN 57	*Allium campanulatum* S. Watson	UCBG 93.1219	UCBG 93.1219	EU096143	EU162712	UC Botanical Garden
HD 07-01	*Allium campanulatum* S. Watson	HD 07-01	—	EU096144		Sequoia National Park, CA, USA
NN 127	*Allium canadense* L. var. *canadense*	NN 127	—	EU096145	EU162713	Gainesville, FL, USA; 29°33′21.04″N, 82°23′25.15″W
NN 58	*Allium cratericola* Eastw.	—	RPBG 82.7	EU096146	EU162714	Regional Parks Botanical Garden
NN 60	*Allium crispum* Greene	—	RPBG 98.5	EU096147	EU162715	Regional Parks Botanical Garden
CS 07-06	*Allium cyathophorum* Bureau & A. Franchet	—	ZG 1969 BL00403	EU096148		Utrecht Botanical Garden
NN 103	*Allium denticulatum* (Ownbey & Aase ex Traub) McNeal	JEPS 91047	—	EU096149	EU162716	Jepson Herbarium
NN 61	*Allium diabolense* (Ownbey & Aase ex Traub) McNeal	—	UCBG 95.0508	EU096150	EU162717	UC Botanical Garden
NN 62	*Allium dichlamydeum* Greene	—	RPBG 92.206	EU096151	EU162718	Regional Parks Botanical Garden
CS 07-12	*Allium dichlamydeum* Greene	CS 07-12	—	EU096152		Mt. Diablo State Park
NN 63	*Allium falcifolium* Hook. & Arn.	—	UCBG 92.0410	EU096153	EU162719	UC Botanical Garden
NN 86	*Allium falcifolium* Hook. & Arn.	NN 86	—	EU096154		UC McLaughlin Reserve, CA, USA; 38°51′36.72″N, 122°25′26.38″W
NN 102	*Allium fimbriatum* S. Watson var. *fimbriatum*	JEPS 97989	—	EU096155	EU162720	Jepson Herbarium
NN 64	*Allium fimbriatum* S. Watson var. *purdyi* (Eastw.) McNeal	—	UCBG 82.1153	EU096156	EU162721	UC Botanical Garden
CS 07-01	*Allium geyeri* S. Watson		1978-BL00074	EU325672	EU 162746	Utrecht University Botanic Garden
NN 65	*Allium haematochiton* S. Watson	—	UCBG 90.0117	EU096157	EU162722	UC Botanical Garden
NN 125	*Allium haematochiton* S. Watson	—	UCBG 87.0043	EU096158	EU162723	UC Botanical Garden
NN 66	*Allium hickmanii* Eastw.	—	UCBG 82.1085	EU096159	EU162724	UC Botanical Garden
NN 104	*Allium hoffmanii* Ownbey ex Traub	JEPS 81076	—	EU096160	EU162725	Jepson Herbarium

(*continued on next page*)

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

Table 1 (*continued*)

NN Col. #	Species and authority	Herbarium voucher no.	Living collection no.	GenBank ITS	GenBank ETS	Coll. locality
NN 100	*Allium howellii* Eastw. var. *howellii*	UC 1561845	—	EU096161	EU325670	UC Herbarium
NN 67	*Allium hyalinum* Curran	—	UCBG 67.1062	EU096162	EU162726	UC Botanical Garden
NN 68	*Allium jepsonii* (Ownbey & Aase ex Traub) S. Denison & McNeal*	—	RPBG 1.075	EU096163	EU162727	Regional Parks Botanical Garden
NN 55	*Allium lemmonii* S. Watson	—	RPBG 92.236	EU096164	EU162728	Regional Parks Botanical Garden
NN 70	*Allium membranaceum* Ownbey ex Traub	—	UCBG 2006.0204	EU096165	EU162729	UC Botanical Garden
NN 71	*Allium obtusum* Lemmon	—	RPBG s.n.	EU096166	EU162730	Regional Parks Botanical Garden
NN 117	*Allium obtusum* Lemmon var. *obtusum*	JEPS 111204	—	EU096167	EU162731	Jepson Herbarium
CS 07-05	*Allium odorum* L.	—		EU096168		Utrecht Botanical Garden
NN 107	*Allium parvum* Kellogg	UC 1585997	—	EU096169	EU325669	UC Herbarium
NN 106	*Allium peninsulare* Lemmon ex Greene var. *peninsulare*	UC 1787693	—	EU096170	EU162732	UC Herbarium
NN 73	*Allium platycaule* S. Watson	—	UCBG 89.1749	EU096171		UC Botanical Garden
NN 74	*Allium platycaule* S. Watson	—	RPBG s.n.	EU096172	EU162733	Regional Parks Botanical Garden
NN 75	*Allium praecox* Brandegee	—	RPBG 96.38	EU096173	EU162734	Regional Parks Botanical Garden
NN 105	*Allium punctum* L. F. Hend.	UC 1561530	—	EU096174	EU162735	UC Herbarium
NN 76	*Allium sanbornii* A. W. Wood	—	RPBG 87.189	EU096175	EU162736	Regional Parks Botanical Garden
NN 77	*Allium sanbornii* A. W. Wood	—	RPBG 87.189	EU096176		Regional Parks Botanical Garden
NN 108	*Allium sanbornii* A. W. Wood var. *sanbornii*	JEPS 100231	—	EU096177	EU325671	Jepson Herbarium
CS 07-08	*Allium serra* McNeal & Ownbey	CS 07-08	—	EU096178	EU162737	Mt. Diablo State Park, CA, USA
NN 135	*Allium sharsmithiae* (Ownbey & Aase ex Traub) McNeal	NN 135	—	EU096179	EU162738	Adobe Canyon, CA, USA; 37°24.517N, 121°24.549W
NN 78	*Allium shevockii* McNeal	—	RPBG 88.168	EU096180	EU162739	Regional Parks Botanical Garden
NN 116	*Allium siskiyouense* Ownbey ex Traub	JEPS 105055	—	EU096181	EU162740	Jepson Herbarium
NN 141	*Allium* sp.	NN 141	—	EU096182	EU162741	Kern County, CA, USA
NN 79	*Allium stellatum* Fras. ex Ker Gawl	—	UCBG 86.0083	EU096183	EU162742	UC Botanical Garden
NN 111	*Allium tribracteatum* Torr.	JEPS 104813	—	EU096184		Jepson Herbarium
NN 123	*Allium tuolumnense* (Ownbey & Aase ex Traub) S. Denison & McNeal*	JEPS 96568	—	EU096185	EU162743	Jepson Herbarium
NN 81	*Allium unifolium* Kellogg	—	UCBG 77.0339	EU096186	EU162744	UC Botanical Garden
CS 07-09	*Allium unifolium* Kellogg	CS 07-09	—	EU096187		Mt. Diablo State Park, CA, USA
NN 83	*Allium validum* S. Watson	—	RPBG s.n.	EU096188	EU162745	Regional Parks Botanical Garden
NN 110	*Allium yosemitense* Eastw.	JEPS 88758	—	EU096189		Jepson Herbarium
NN 49	*Dichelostemma capitatum* (Benth.) Keator spp. *capitatum*	NN 49	—	EU096190		Tilden Regional Park, CA, USA; 37°54′14.32″ N, 122°15′16.95″ W
NN 52	*Dichelostemma congestum* Kunth	—	UCBG 98.0436	EU096191		UC Botanical Garden
NN 142	*Dichelostemma ida-maia* Greene	—	RPBG 88.61	EU096192		Regional Parks Botanical Garden
NN 124	*Dichelostemma multiflorum* A. Heller	—	RPBG 77.058	EU096193		Regional Parks Botanical Garden
NN 144	*Dichelostemma volubilis* (Kellogg) A. Heller	NN 144	—	EU096194		Stebbins Cold Canyon Reserve, CA, USA; 38°50.00′N, 122°10.28′W
NN 126	*Nothoscordum gracile* (Aiton) Stern	NN 126	—	EU096195		Tilden Regional Park, CA, USA; 37.91°N, 122.28°W

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

Table 1 (*continued*)

GenBank Accessions for previously published *Allium* taxa: *Allium aflatunense* (AF037615), *A. albidum* (AJ411954), *A. altaicum* (AJ412750), *A. ampeloprasum* (AJ411888), *A. amplectens* (AF055097), *A. anisopodium* (AJ411847), *A. asarense* (AJ411937), *A. atropurpureum* (AF037616), *A. atrosanguineum* (AJ411864), *A. atroviolaceum* (AJ411884), *A. beesianum* (AJ411860), *A. bidentatum* (AJ411861), *A. bolanderi* (AF055101), *A. brevidens* (AJ412721), *A. brevistylum* (AJ412763), *A. caeruleum* (AJ412729), *A. caeruleum* (AJ412729), *A. caesium* (AJ412731), *A. caesium* (AJ412731), *A. cepa* (AJ411944), *A. cernuum* (AJ250289), *A. chamaemoly* (AF055109), *A. chamarense* (AJ411957), *A. chinense* (AJ411848), *A. christophii* (AF037610), *A. clathratum* (AJ411855), *A. condensatum* (AJ412752), *A. crystallinum* (AJ412724), *A. cupanii* (AJ412737), *A. cyaneum* (AJ411880), *A. daghestanicum* (AJ411850), *A. darwasicum* (AF037617), *A. dentigerum* (AJ411958), *A. dichlamydeum* (AF055098), *A. dregeanum* (AJ411962), *A. drepanophyllum* (AJ411854), *A. drobovii* (AJ411895), *A. drummondii* (AJ411908), *A. elegans* (AJ412730), *A. eremoprasum* (AJ412726), *A. farctum* (AM492184), *A. fedschenkoanum* (AJ411916), *A. fetisowii* (AF037619), *A. filidens* (AJ412723), *A. filidentiforme* (AJ412722), *A. fistulosum* (AM418372), *A. flavum var. minus* (AJ411926), *A. galanthum* (AJ411905), *A. giganteum* (AF037607), *A. gilgiticum* (AJ412762), *A. glandulosum* (AJ412746), *A. goodingii* (AF055095), *A. griffithianum* (AJ411862), *A. gunibicum* (AJ411890), *A. haneltii* (AJ412725), *A. heldreichii* (AY427539), *A. hirtifolium* (AF037612), *A. hookeri* (AJ250297), *A. hyalinum* (AF055099), *A. hymenorrhizum* (AJ411879), *A. inaequale* (AJ412735), *A. insubricum* (AJ250291), *A. iranicum* (AJ411961), *A. jodanthum* (AJ411902), *A. kaschianum* (AJ412754), *A. kingdonii* (AJ250286), *A. komarovianum* (AJ412760), *A. kopetdagense* (AJ411950), *A. kunthianum* (AJ412734), *A. kuramense* (AJ411868), *A. lineare* (AJ411951), *A. litvinovii* (AJ412727), *A. macleanii* (AF037608), *A. macrostemon* (AJ412738), *A. mairei* (AJ250298), *A. margaritae* (AJ412732), *A. melanantherum* (AJ412739), *A. minorensis* (AJ412748), *A. moly* (AJ412743), *A. monadelphum* (AJ411955), *A. monanthum* (AJ412745), *A. mongolicum* (AJ411883), *A. moschathum* (AJ411872), *A. neriniflorum* (AJ411920), *A. obliquum* (AJ412753), *A. ochroleucum* (AJ411856), *A. oreophilum* (AJ411931), *A. oreoprasoides* (AJ411896), *A. oreoprasum* (AJ411933), *A. oschaninii* (AJ411940), *A. ovalifolium* (AJ411882), *A. pamiricum* (AJ412736), *A. paniculatum* (AJ411949), *A. paradoxum* (AJ412741), *A. parvulum* (AJ412720), *A. petraeum* (AM418363), *A. platyspathum* var. *amblyophyllum* (AJ411875), *A. polyrhizum* (AJ250296), *A. praemixtum* (AJ411873), *A. protensum* (AF037609), *A. przewalskianum* (AJ411852), *A. pskemense* (AJ411907), *A. ramosum* (AJ250295), *A. roseum* (AF055105), *A. roylei* (AJ411945), *A. rupestre* (AJ412733), *A. sanbornii* (AF055103), *A. sativum* (AJ411901), *A. saxatile* (AY427545), *A. scabriscapum* (AJ411881), *A. schmitzii* (AJ412761), *A. schoenoprasoides* (AJ412728), *A. schoenoprasoides* (AJ412728), *A. schoenoprasum* (AJ411836), *A. schubertii* (AF037611), *A. scorodoprasum* (AJ412713), *A. semenovi* (AJ411897), *A. senescens* (AY427548), *A. setifolium* (AJ411898), *A. siculum* (AJ250299), *A. sikkimense* (AJ411885), *A. sordidiflorum* (AJ411899), *A. spirale* (AJ411833), *A. splendens* (AJ411927), *A. stellatum* (AF055102), *A. suaveolens* (AJ411874), *A. subangulatum* (AJ411870), *A. subhirsutum* (AJ411912), *A. sulphureum* (AJ412759), *A. talassicun* (AJ411865), *A. tanguticum* (AJ411893), *A. tenuissimum* (AJ411846), *A. teretifolium* (AJ411886), *A. thunbergii* (AJ411849), *A. togashii* (AJ411843), *A. trachyscordum* (AJ411857), *A. tricoccum* (AJ411917), *A. triquetrum* (AJ412742), *A. tuberosum* (AJ250293), *A. turkestanicum* (AJ412718), *A. ubsicolum* (AJ411960), *A. umbilicatum* (AJ412719), *A. unifolium* (AF055100), *A. ursinum* (AJ412744), *A. validum* (AF055096), *A. vavilovii* (AJ411840), *A. vodopjanovae* (AJ411845), *A. vodopjanovae* (AJ411942), *A. vodopjanovae* var. *czemalense* (AJ311868), *A. wallichii* (AJ250294), *A. weschniakowii* (AJ411946), *A. xiphopetalum* (AJ411858), *A. zebdanense* (AF055107), *A. zebdanense* (AY427552), *Ipheion uniflorum* (AJ412715), *Nectaroscordum* (*Allium*) *bulgaricum* (AJ412747), *Nothoscordum bivalve* (AJ250301), *Tulbaghia fragrans* (AJ250300).

sequence, saving all trees at each step (MulTrees) with ACCTRAN character state optimization. All characters were treated as unordered and equally weighted. Bootstrap analyses (Felsenstein, 1985) were conducted to assess topological support under the following conditions for each dataset: 1000 bootstrap replicates with simple taxon addition, TBR branch swapping, and MaxTrees set to 100 for the global ITS dataset and 1000 for the combined ITS/ETS dataset.

Bayesian phylogenetic analyses were performed with MrBayes v. 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Each dataset was assigned its own model of nucleotide substitution, as determined by a hierarchical likelihood ratio test and Akaike information criterion (AIC) in Modeltest v. 3.06 (Posada and Crandall, 1998). Posterior probabilities of the generated trees were approximated using a Markov chain Monte Carlo (MCMC) algorithm with four incrementally heated chains ($T = 0.2$) for 5,000,000 generations and sampling trees every 100 generations. Two independent runs were conducted for each dataset simultaneously, the default setting in MrBayes v. 3.1.2. Following completion, the sampled trees from each analysis were plotted against their log-likelihood score to identify the point where log-likelihood scores reached a maximum value. All trees prior to this point were discarded as the burn-in phase, all post-burn-in trees from each run were pooled, and a 50% majority-rule consensus tree was calculated to obtain a topology with average branch lengths as well as posterior probabilities for all resolved nodes.

2.5. Character state reconstruction

To reconstruct ancestral character states for species localized on serpentine soils, we used the combined ITS and ETS dataset (with multiple representatives of a species removed) to find the maximum likelihood (ML) tree, using the best-fit model of nucleotide substitution, the General Time Reversible model (GTR) with a proportion of invariable sites (I) and a gamma distribution parameter (G), for 1,000,000 generations in GARLI (Zwickl, 2006). We imported a binary character matrix (serpentine vs. non-serpentine) into Mesquite (Maddison and Maddison, 2006) and performed an ancestral character state reconstruction using the ML tree with branch lengths from GARLI and the ML reconstruction criteria native to Mesquite.

3. Results

3.1. Sampling

We sampled *Allium* species from wild collections, botanical gardens, and herbarium specimens (Table 1). Material from botanical gardens for the most part was correctly identified, but several specimens were mislabeled. We corrected the identification using the Jepson Manual (Hick-

man, 1993) in combination with molecular data that indicated potential hybrid origin or major differences from expected. Where possible, two different exemplars from different living collections were used for any species in order to confirm identities. Identical sequences for different accessions were submitted to GenBank, but only one was included in the analyses.

DNA was easily extracted and both ITS and ETS amplified from fresh material, but herbarium specimens required several rounds of amplifications using different primer sets to optimize amplification of PCR product from degraded material. In total, we were able to include 39 out of 48 species that occur natively in the California Floristic Province. Taxonomic determination of varieties can be difficult, often relying on plastic or homoplasious characters such as color and bulb shape (Fig. 1; M). Therefore, we sampled multiple accessions for cryptic species. We also included recognized varieties when possible in order to determine if any genetic differences exist in the targeted genes and to investigate any potential intraspecific variation in the targeted gene regions (ITS and in some cases ETS). Our data show that there were detectable genetic differences among and within the varieties of species such as *A. fimbriatum* (var. *fimbriatum*, var. *purdyi* [Fig. 1; O]), *A. sanbornii* (var. *sanbornii* [Fig. 1; D]), *A. obtusum* (var. *obtusum*), *A. atrorubens* (var. *atrorubens*, var. *cristatum*), and *A. bolanderi* (var. *bolanderi* [Fig. 1; L]). In all cases in which multiple exemplars were sampled for a named species or multiple varieties were sampled within a species, the species as defined was found to be monophyletic with the exception of *A. sanbornii* which forms a clade containing the species *A. jepsonii*. For the ITS region, genetic variation among varieties, however, was not significantly different than genetic variation found between multiple exemplars of a single species.

3.2. Molecular datasets

Two different datasets were generated: dataset 1 is composed of only ITS with a worldwide sampling of *Allium* species. Dataset 2 is a taxonomic subset of dataset 1 with a focus on subgenus *Amerallium* with the ITS data realigned where necessary and the addition of ETS for the majority of the species. Dataset 2 (ITS + ETS) includes all species included in the monophyletic subgenus *Amerallium* based on the analysis of dataset 1, plus closely related *A. bulgaricum* and *A. siculum*. The same outgroups were used in both datasets. ITS sequences ranged from 615 to 720 bp and ETS sequences ranged from 397 to 485 bp (ITS for *Allium geyeri* and ETS for *Allium yosemitense* were incomplete and are not included in length calculations). The global ITS alignment (dataset 1) was 826 bp in length and resulted in 163 constant characters and 663 variable characters of which 579 were parsimony-informative. For ETS, the aligned sequences produced a matrix 483 bp in length of which 165 characters were constant, 318 variable, and 245 parsimony-informative. The combined ITS/ETS aligned dataset (dataset 2) was 1253 bp in length, with 389 characters constant, 864 variable, and 710 parsimony-informative.

3.3. Phylogenetic reconstruction

We generated a phylogenetic hypothesis for each of the datasets described above (Figs. 2 and 3). All datasets were analyzed using Bayesian and parsimony methods. The best-fit model of nucleotide evolution for both datasets and implemented in the Bayesian analyses was GTR+I+G. Regardless of taxon sampling or analysis method, the phylograms produced were very similar with high statistical support for grouping of species. Our analyses of the global ITS dataset showed no significant difference in topology between the Bayesian (Fig. 2) and maximum parsimony consensus trees (data not shown). ETS combined with ITS provided some additional resolution for section *Amerallium*. All major relationships were the same in both the Bayesian and parsimony consensus trees and were well supported by Bayesian posterior probabilities and parsimony bootstrap values. Because Bayesian and parsimony topologies were similar, we chose to show one of the 1296 most parsimonious trees (Fig. 3).

3.4. Character state reconstruction

Character state reconstruction resulted in mostly unambiguous ancestral states for occurrence on serpentine (Fig. 4). The North American ancestor most likely did not occur on serpentine, however the *Lophioprason* ancestor, in addition to being reconstructed as having $n = 7$ and occurring in the California Floristic Province (data not shown), has a 45% chance of having occurred on serpentine based on likelihood reconstruction. If that ancestor did occur on serpentine, then adaptation to serpentine would have been independently lost six times and gained five times. If the *Lophioprason* ancestor did not occur on serpentine soils, which is slightly more likely, then we record six independent gains and five losses. Both situations are equally parsimonious; however, when branch lengths are considered in the ancestral reconstruction (Fig. 4), the latter scenario is favored.

4. Discussion

4.1. Global phylogeny

The overall phylogeny for the genus *Allium* (Fig. 2) shows a number of well-supported clades and subclades therein that, for the most part, reflect groups that were classified into sections based on traditional taxonomy (Traub, 1968, 1972) or have been considered in the new classification of Friesen et al. (2006). The native North American alliums (Fig. 2; "North American *Amerallium*") are found to be monophyletic with bootstrap proportion (bp) of 90% and Bayesian posterior probability (pp) of

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

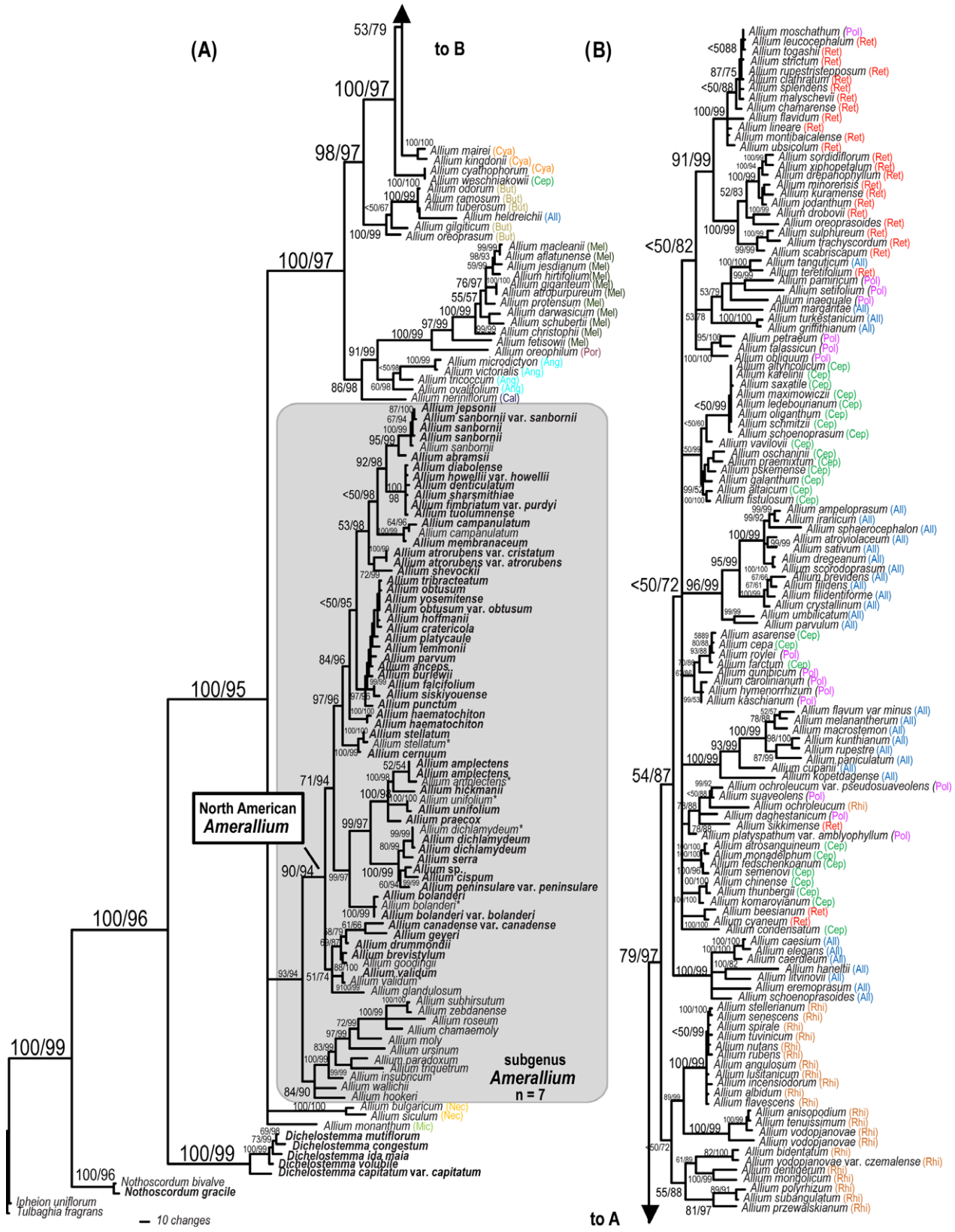

Fig. 2. A Bayesian tree showing the relationships within the genus *Allium*. Values on branches represent parsimony bootstrap and Bayesian posterior probabilities, respectively. Sequences deposited to GenBank from this study are listed in bold. Colored three letter codes in parentheses represent the new subgeneric classification according to Friesen et al., 2006 and Dubouzet and Shinoda, 1999 (with asterisk*), except for subgenus *Amerallium* which is highlighted gray; All = *Allium*, Ang = *Anguinum*, But = *Butomissa*, Cal = *Caloscordum*, Cep = *Cepa*, Cya = *Cyathophora*, Mel = *Melanocrommyum*, Mic = *Microscordum*, Nec = *Nectaroscordum*, Pol = *Polyprason*, Por = *Porphyroprason*, Ret = *Reticulatobulbosa*, Rhi = *Rhizirideum*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

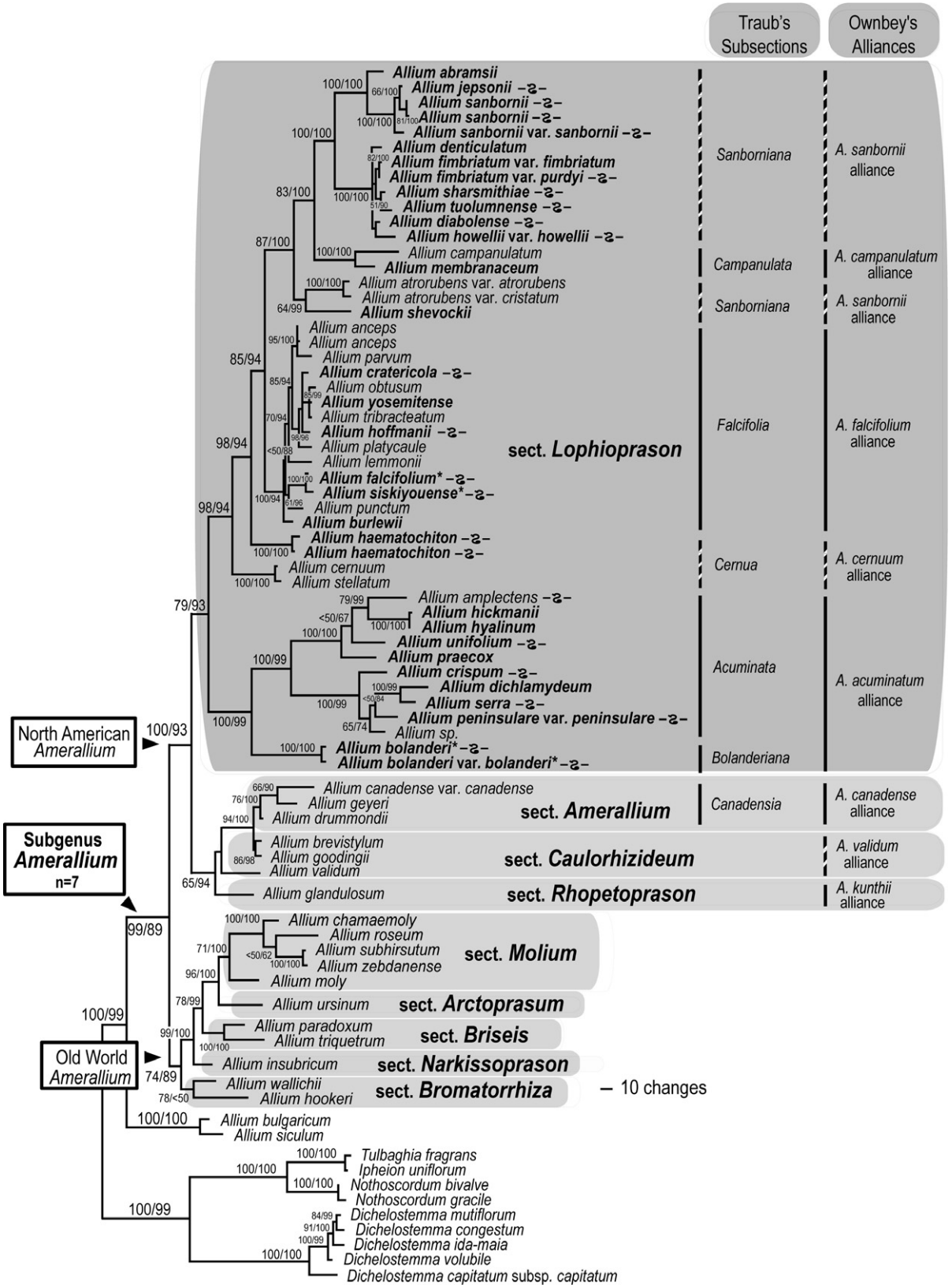

Fig. 3. One of 1296 most parsimonious trees (MPT) showing the relationship of subgenus *Amerallium* using a combined ITS and ETS dataset. Values on branches represent bootstrap proportions and Bayesian posterior probabilities, respectively. Length = 2774, CI = 0.526, RI = 0.807. Endemic California species are listed in bold. * = species endemic to both northern California and southern Oregon within the same geological boundary. **–ƨ–** = species that are adapted to living on serpentine soils; this symbol only applies to the species level, as there are some varieties that are not serpentine-adapted. Subsections and alliances recognized by Traub and Ownbey (respectively) are indicated, demonstrating the strong correlation between the phylogenetic results and previous taxonomic groupings. Vertical slashed lines represent groupings that are not monophyletic.

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

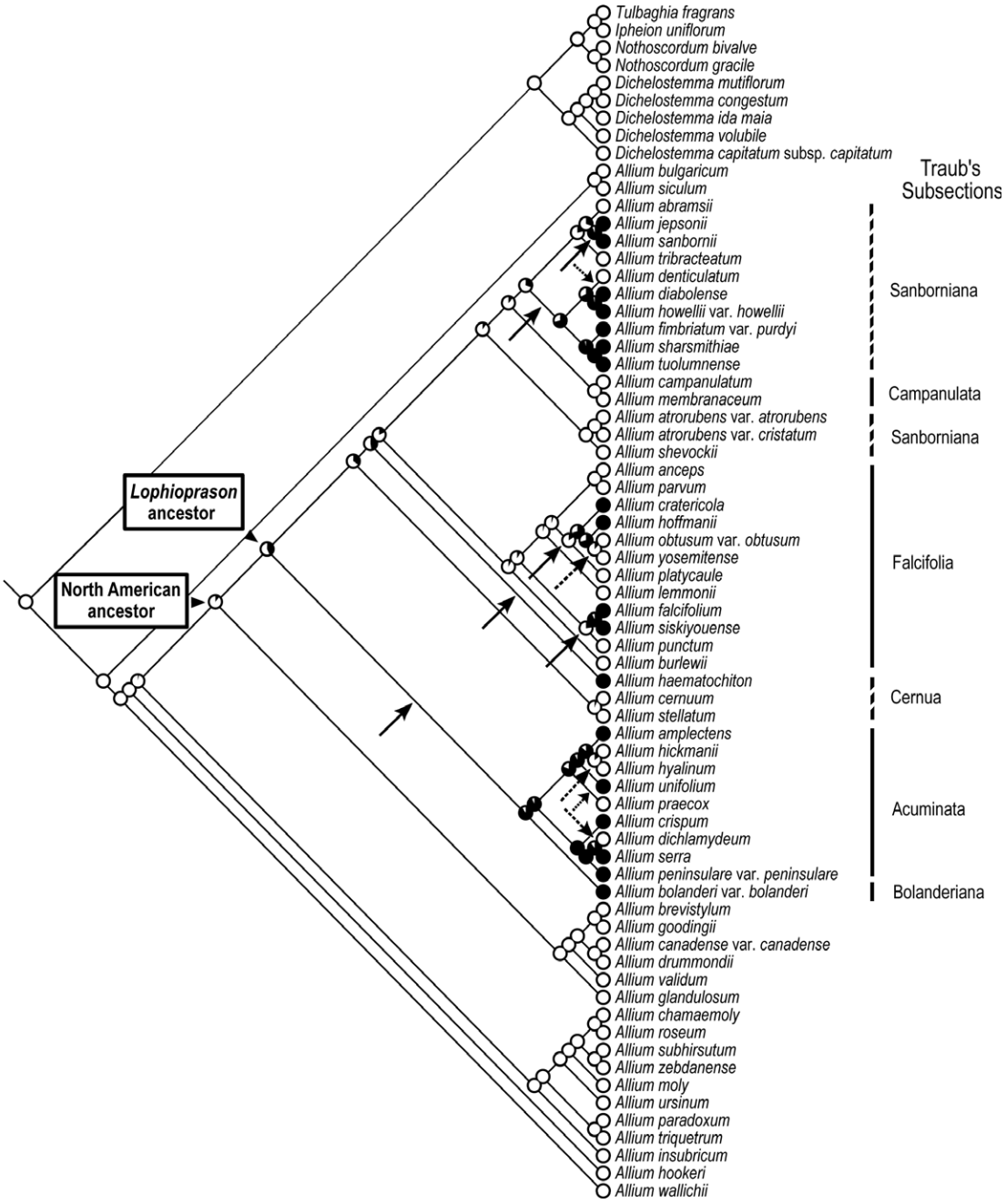

Fig. 4. Character state reconstruction of serpentine soil adaptation in section *Lophioprason*. The maximum likelihood probability of adaptation is presented at each node by a pie graph, where black is serpentine and white is non-serpentine. Gains (solid arrows) and losses (broken arrows) are shown based on the *Lophioprason* ancestor being non-serpentine. Subsections recognized by Traub are indicated. Vertical slashed lines represent groupings that are not monophyletic.

94%. They are included within the larger section *Amerallium* (bp = 93/pp = 94), reflecting also the $n = 7$ base chromosome number as reported previously (Dubouzet and Shinoda, 1999; Friesen et al., 2006). *Allium tricoccum* ($n = 8$) represents a special case where although considered to be a North American species, it falls outside of the $n = 7$ section *Amerallium*. As found in Klass and Friesen (2002), *A. tricoccum* is sister to the Old World species *A. ovalifolium* ($n = 8$, subgenus *Anguinum*), indicating dispersal to North America that occurred independently of the dispersal event leading to the main North American clade (Fig. 2). Similar to the history of the holarctic species *A. schoenoprasum* (a post-glacial migrant, Dubouzet and Shinoda, 1999), *A. tricoccum* may have been a recent introduction to North America. It is important to note that we based this phylogenetic inference on only one sequence from *A. tricoccum* that was obtained from GenBank. Additional sequences from multiple exemplars and from a variety of genomic sources will be necessary to confirm this hypothesis of dispersal.

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

While resolution within *Amerallium* was good, our single-marker phylogeny was not able to resolve some of the more recent relationships among Old World clades (Fig. 2; note polytomy and/or non-monophyly of subgenera *Allium*, *Cepa*, *Polyprason*, and *Reticulatobulbosa*). Additional sampling of rapidly evolving chloroplast or low-copy nuclear regions will be necessary to further investigate relationships in this region of the phylogeny.

4.2. Taxonomic considerations

4.2.1. Subgenus Amerallium

We next analyzed a subset of the global ITS dataset combined with an ETS dataset, which focused on subgenus *Amerallium* (Fig. 3). Previous traditional classifications have placed all *Allium* species with chromosome base numbers of $n = 7$ into subgenus *Amerallium*, which contains both Old World and North American species, a placement that is supported by our molecular phylogenetic results (Fig. 3; "Subgenus *Amerallium*"). The North American taxa were monophyletic and sister to the Old World taxa (bp = 99/pp = 89) with moderate statistical support for the monophyly of the Old World clade (Fig. 3; "Old World *Amerallium*"; bp = 74/pp = 89) and high support for the monophyly of the North American clade (Fig. 3; "North American *Amerallium*"; bp = 100/pp = 93). In the Old World *Amerallium* clade, the monophyletic section *Bromatorrhiza* Ekberg (=*A. wallichii* and *A. hookeri*) was sister to a clade containing all other sections: *Narkissoprason* Kam. (=*A. insubricum*), *Briseis* (Salisb.) Stearn (=*A. paradoxum* and *A. triquetrum*), *Arctoprasum* Kirschl. (=*A. ursinum*), *Molium* (=*A. moly*, *A. subhirsutum*, *A. zebdanense*, *A. roseum*, and *A. chamaemoly*). Sections *Briseis* and *Molium* were monophyletic, thus reflecting Friesen et al. (2006). *Allium chamaemoly* L. has been previously placed in subgenus *Chamaeprason* Herm., but is treated here as part of section *Molium* as was suggested by de Wilde-Duyfjes (1976) and was also found in the molecular analysis of Dubouzet and Shinoda (1999). *Allium ursinum*, while sister to the monophyletic *Molium*, is maintained as a separate section due to various unique characteristics that differentiate section *Arctoprasum* (=hypogeal seed germination and unique seedling and leaf morphology, with the upper and lower leaf margins being reversed anatomically). We should note here that these sequences were obtained from GenBank, and sampling for Old World *Amerallium* remains incomplete.

Within the well-supported North American *Amerallium* clade, species were divided into several subclades corresponding to sections *Lophioprason* (bp = 79/pp = 93) and *Amerallium* + *Caulorhizideum* + *Rhopetoprason* (Fig. 3; bp = 65/pp = 94). Our sampling focus was on California species, all of which belong to the monophyletic section *Lophioprason*, which we found to be monophyletic with moderate to good support (bp = 79/pp = 93). The sister relationship of this clade to the North American sections *Amerallium* + *Caulorhizideum* + *Rhopetoprason* is well supported (bp = 100/pp = 93).

4.2.2. Section Lophioprason

The intrageneric taxonomy of North American alliums have been investigated thoroughly by several authors, two of the most notable being Ownbey and Traub and most recently McNeal. These authors used a variety of traditional characters such as leaf morphology (sickle [Fig. 1; A], round, channeled), bulb scale (Fig. 1; N), lacticifers, and ovary crests for monographic identification of species, as well as for proposing generic-level taxonomy that included section and subsection affiliations. Ownbey traditionally grouped all of the North American *Allium* species into 10 alliances, eight of which belonged to subgenus *Amerallium* and two of which (*A. schoenoprasum* and *A. tricoccum* alliances) were considered to be unrelated to the North American *Amerallium* group or to each other (Saghir et al., 1966). The latter two contain only the nominal species and are likely the result of separate and recent introductions to the New World potentially as cultivated species that became naturalized (*A. schoenoprasum* = cultivated chives). Those species belonging to *Amerallium* were reported to have a base chromosome number of $n = 7$, while the two monotypic alliances have base chromosome number of $n = 8$. Their phylogenetic positions (see Fig. 2) clearly place them as outside the New World *Amerallium* clade.

Within the eight alliances considered to be part of *Amerallium*, five would have been included in what was traditionally referred to as section *Lophioprason*. Ownbey's alliance groupings are not valid under the International Code of Botanical Nomenclature (Saghir et al., 1966; McNeal, 1992) but are useful to assess the level of diversity and to begin defining taxonomic affinities encompassed in *Amerallium* and in *Lophioprason*. Traub (1968, 1972) elevated these alliances into eight subsections, six of which are included in section *Lophioprason*. The subsectional groupings, while obtaining new status under the ICBN, did not change dramatically and primarily used Ownbey's characters for group designation. These changes, although formal, have been largely ignored by North American *Allium* workers perhaps due to the little evidence supporting their classification (McNeal, 1992). McNeal recognized that subsectional classification was premature and awaited further evidence. For clarity, henceforth we use Traub's classification to discuss relevant clades.

Our molecular phylogeny for *Allium* section *Lophioprason* supports the subsectional status of the various morphological groupings. The six subsections or five alliances in section *Lophioprason* correspond relatively well to the clades supported in this analysis (Fig. 3). Two main clades are formed within *Lophioprason*, the first comprised of subsections *Bolanderiana* + *Acuminata* (bp = 100/pp = 99) and the second containing the remaining subsections (*Cernua, Falcifolia, Sanborniana*, and *Campanulata*; bp = 98/pp = 94). The only difference between Traub's subsections

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

and Ownbey's alliances within *Lophioprason* is Traub's additional segregation of *Allium bolanderi* (Fig. 1; L, M) and its varieties into its own subsection (=*Bolanderiana*). Ownbey considered *A. bolanderi* to be part of the *A. acuminatum* alliance (see Fig. 3). In either system for the most part, the proposed subsections or alliances are found to represent monophyletic groups. The uniqueness of *A. bolanderi* and its strong support as a clade (bp = 100/pp = 99) placed sister to all other species in subsection *Acuminata* (bp = 100/pp = 99) argue for maintaining this subsection as unique.

Subsection *Cernua* consists of several species, including *A. haematochiton, A. cernuum*, and *A. stellatum*. *Allium cernuum* and *A. stellatum* were found to be sister species (bp = 100/pp = 100) but were not placed in a clade with *A. haematochiton* (Fig. 3) regardless of the method of analysis. Rather, they were placed as sister to the remaining species in the clade comprised of *A. haematochiton* plus members of subsections *Sanborniana*, *Campanulata*, and *Falcifolia*. *Allium haematochiton* may deserve its own subsectional status, as in both analyses it is resolved as sister to the clade containing *Sanborniana*, *Campanulata*, and *Falcifolia* taxa (Fig. 3; bp = 98/pp = 94). Subsection *Falcifolia* was strongly supported as monophyletic. The grouping of species in this subsection was traditionally well supported by their distinctive sickle-shaped leaf morphology.

Most species of subsection *Sanborniana* grouped together into a clade with good support (Fig. 3; bp = 87/pp = 100), however this clade also contained members of subsection *Campanulata* (*A. campanulatum, A. membranaceum*) making subsection *Sanborniana* non-monophyletic. *Allium shevockii* and *A. atrorubens*, members of subsection *Sanborniana*, are sister to one another and together are sister to the *Sanborniana* + *Campanulata* clades (see Fig. 3) with moderate-to-strong support (bp = 87/pp = 100). Subsection *Campanulata* was found to be monophyletic, but the clade is placed as sister to all included species of *Sanborniana* minus *Allium atrorubens* and *A. shevockii* (Fig. 3).

Overall, most of the groupings made by using traditional morphology were well supported by molecular evidence. However, our current lack of certainty regarding the placement of *A. shevockii* (Fig. 1; G), *A. atrorubens*, and *A. hematochiton* prevents us from suggesting their classification at this time. Additional phylogenetic efforts are needed to obtain solid molecular evidence that supports (or rejects) suggested morphological synapomorphies prior to revising section *Lophioprason*.

4.2.3. Sections Amerallium, Caulorhizideum, and Rhopetoprason

While these sections were not a focus of our analysis, it is important to relate the taxonomy of sections *Amerallium*, *Caulorhizideum*, and *Rhopetoprason* within the context of the discussion above. These three subsections corresponded with Ownbey's three *Allium canadense*, *Allium validum*, and *Allium kunthii* alliances. Within the monophyletic section *Amerallium* Traub recognized two subsection, *Canadensia* and *Mexicana*. Our sampling only include subsection *Canadensia*. Sections *Caulorhizideum* and *Rhopetoprason* were not divided further into subsections. Section *Caulorhizideum* is found in our analysis to be non-monophyletic because of the position of *A. validum* as sister to a clade containing the remaining *Caulorhizideum* + *Amerallium* (see Fig. 3). *Allium glandulosum*, placed by Ownbey in the *Allium kunthii* alliance but most recently considered part of section *Rhopetoprason* (Friesen et al., 2006), is sister to all other species in the *Amerallium* + *Caulorhizideum* clade. It is distinguished by a long branch indicating significant genetic distance (Fig. 3). Additional representatives of this alliance (or section, if considered *Rhopetoprason*) will be needed to test its monophyly and unique position as sister to *Caulorhizideum*. It is worth mentioning here that we did not have ETS data for most of the taxa in the three sections just discussed. Future studies should include a thorough sampling of taxa using a multi-gene approach to resolve the relationships within these clades.

4.3. Natural History of Amerallium section Lophiprason

Previous studies suggest that the subgenus *Amerallium* had its origins in Asia and spread to North America through the Bering and North Atlantic land bridges (Hanelt et al., 1992), although later studies note that such a hypothesis does not explain the separation of Mediterranean and North American centers of diversity and the comparable dearth of species in the intervening continental Asia (Dubouzet and Shinoda, 1999). In his treatise on the Madrean–Tethyan sclerophyll vegetation, Axelrod (1975) proposed that gene flow may have continued between North America and the Mediterranean in the early Oligocene (34 mya) until the end of the Paleogene (25 mya), thus enabling *Amerallium* to remain connected as a clade with eventual separation into North American and Mediterranean clades. Previous authors emphasized that detailed sampling of North American taxa would be necessary to elucidate the role of geographic formations and to discern possible biogeographic patterns of *Amerallium*, especially with respect to the differences between the Bering Strait migration v. Madrean–Tethyan gene flow and expected subsequent patterns for North American distributions.

Based on our results, the first branch within the American alliums separates a clade of species native to mid-western and southwestern United States from section *Lophioprason* containing Western North America (and especially California endemic) species. While the North American clade as a whole is clearly monophyletic (bp = 100/pp = 93) and suggests a single colonization of the New World within the *Amerallium* lineage, this early separation implies that two separate biogeographic patterns led to *Allium* diversification in North America. One clade remained in the central plains region, extending into

Canada and eastern (Gulf Coast) Mexico along the central and eastern plateaus of North America without crossing the Rocky Mountains (with few exceptions, e.g., *A. validum*). We do not currently have enough sampling outside of *Lophioprason* to evaluate the biogeographic patterns in the *Caulorhizideum* clade, however preliminary analysis indicates that this clade is focused around an ancestral center of diversity in Texas with dispersal events leading to species with distributions across the mid-west and eastern regions of North America.

In contrast, section *Lophioprason* comprises *Allium* species that are native to California or restricted to western North America in their distributions. The only exceptions are *A. cernuum* and *A. stellatum*, the former distributed from Canada to Mexico but excluding Nevada and California and the latter found throughout the central plains region east of the Rocky Mountains and (primarily) west of the Mississippi River drainage. These taxa were previously placed in subsection *Cernua* along with *A. haematochiton*, a Californian endemic that is not included in the Cernua clade in this analysis. Our Cernua clade (see Fig. 3) is well supported as sister to a larger clade containing both California natives and other northwestern species placed taxonomically into subsections *Sanborniana*, *Campanulata*, and *Falcifolia*.

Californian endemics are found in both of the major clades of section *Lophioprason* and are scattered throughout all alliances/subsections included in the subgenus with the exclusion of subsection *Cernua*. Only one clade of subsection *Sanborniana* is entirely comprised of species endemic to California; within this clade, one species (*A. munzii*, data not shown) is federally listed as endangered and one other species (*A. yosemitense*) is listed as rare/threatened. While the California alliums are not a monophyletic lineage according to the phylogenetic hypothesis presented here (Fig. 3; bold), ancestral character state reconstruction using a maximum likelihood criterion (Fig. 4), based on the current sampling, predicts that the ancestor of section *Lophioprason* was native to California (75%, data not shown) and that migrations from California populations enabled the expansion of the genus to include the northwestern regions of North America. Based on this analysis, the "escape from California" occurred multiple times, and in some cases occurred within serpentine-adapted species (see below). It is noteworthy that species in the *Bolanderiana/Acuminata* clade and one subclade in *Sanborniana* remain in California for the most part (except *A. amplectens*) while species in the second clade of *Sanborniana* and in the *Campanulata/Falcifolia* clade seem to have diversified geographically. Because the Cernua clade is sister to the clade containing subsections *Sanborniana/Campanulata/Falcifolia*, it is possible that the ancestor of this group was distributed more broadly and subsequent reductions in distribution led to the *Sanborniana, Campanulata*, and *Falcifolia* California endemics. While possible, this would mean a total of six and possibly more (depending on parameters for reconstruction) reductions in distribution, most likely caused by 'budding off' of isolated populations or ecological factors (eg. adaptation to serpentine soils) that enabled expansion of a population with limited continued gene flow. Without a better understanding of the gene flow among these species, it is difficult to determine the speciation mechanisms within these groups. Ongoing studies are focusing on these areas.

4.4. Serpentine soils and Allium evolution

Regardless of how *Allium* finally arrived at the western coast of North America, the early species appear to have proliferated and diverged on many types of soil, thriving in the well-drained soils and the Mediterranean climate. One soil type that seems to be important to the evolutionary history of *Allium* is a type of ultramafic soil called serpentine. Serpentine soils occur in patches along the Pacific States and, because of its toxicity, very few plants can withstand the conditions that it offers for growth. Many *Allium* species demonstrate an ability to grow on serpentine, some with tolerance and others endemic to serpentine outcrops indicating adaptation to this soil type. This edaphic endemism seems to drive speciation in *Allium* and could be one of the factors contributing to the greater diversity of *Allium* in California and the surrounding western states than anywhere else in Northern America (Kruckeberg, 1984; Hickman, 1993). There appear to have been multiple introductions to serpentine soil (Fig. 3). Some clades contain both serpentine and non-serpentine species, where only one clade in subsection *Sanborniana* contains only serpentine endemics.

Adaptation to serpentine soils is ongoing, as evidenced by species that can survive on both serpentine and non-serpentine soils, such as *A. amplectens* (Fig. 1; J) and *A. haematochiton* (Fig. 1; H, I), as well as certain species with varieties that are serpentine endemics. Like other plants that can survive on both types of soils, *A. haematochiton* exhibits a different flower morphology (pers. obs.; Wright et al., 2006) when growing on serpentine. Further work will be necessary to determine if local adaptation in this species has occurred on this specialized substrate.

Assuming that the ancestor of section *Lophioprason* was not a serpentine endemic (Fig. 4), serpentine endemism is gained six times and lost five times throughout the course of evolution of the western North American alliums. Loss of the serpentine habit occurs in *Allium denticulatum* of subsection *Sanborniana* and in the ancestor of *A. obtusum* and *A. yosemitense* of subsection *Falcifolia*. Provided that the ancestor of the *Acuminata/Bolanderiana* clade was a serpentine endemic (Fig. 4), a total of three losses occur independently in the *Acuminata/Bolanderiana* clade: one in *A. dichlamydeum*, once in the lineage leading to *A. hickmanii* and *A. hyalinum*, and once in *A. praecox*. This assumes that serpentine endemism was gained in the ancestor of the *Bolanderiana/Acuminata* lineage, as indicated by our likelihood reconstruction.

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006

5. Conclusion

We present a phylogenetic study of North American *Allium* focusing on section *Lophioprason*. These results represent a first attempt at placing *Lophioprason* in the context of the global distribution of the holarctic genus *Allium* and at initiating a more detailed study on the biogeography of subgenus *Amerallium* in North America. A phylogenetic analysis including the remaining unsampled California species in combination with other North American taxa will be necessary to provide a complete phylogeny of North American *Allium*. Analyses that include a broad sampling of both Old World and New World species are essential to understand these relationships, to develop a cohesive phylogeny of *Amerallium*, and to define its relationship to species currently placed within *Lophioprason*. Updates to the infrageneric taxonomy await this increased sampling and additional multigene analyses.

Acknowledgments

We thank Hillary Cooper for designing *Allium*-specific ETS primers, Irene Liao for expert assistance in data acquisition and tireless persistence with ETS sequencing, Jason Stajich for assistance with the UCB cluster, and Patrick O'Grady (UCB) and Catherine Koehler (McLaughlin Reserve) for assistance in field collections. We also acknowledge the following for providing specimens used in this study: The University of California Botanical Garden at Berkeley (especially Holly Forbes), The University and Jepson Herbaria (especially Andrew Doran), the Regional Parks Botanical Garden, and Rancho Santa Ana Botanic Garden. Funding was generously provided by the Lawrence R. Heckard Fund of the Jepson Herbarium.

References

Axelrod, D.I., 1975. Evolution and biogeography of Madrean–Tethyan sclerophyll vegetation. Annals of the Missouri Botanical Garden 62, 280–334.

Baldwin, B.G., Markos, S., 1998. Phylogenetic utility of the external transcribes spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). Molecular Phylogenetics and Evolution 10, 449–463.

Baum, D.A., Small, R.L., Wendel, J.F., 1998. Biogeography and floral evolution of *Baobabs* (Adansonia, Bombacaceae) as inferred from multiple data sets. Systematic Biology 47, 181–207.

Blattner, F.R., Weising, K., Banfer, G., Maschwitz, U., Fiala, B., 2001. Molecular analysis of phylogenetic relationships among myrmecophytic *Macaranga* species (Euphorbiaceae). Molecular Phylogenetics and Evolution 19, 331–334.

Clausen, J., Keck, D., Hiesey, W.M., 1948. Experimental Studies on the Nature of Species. III. Environmental Responses of Climatic Races of Achillea. Carnegie Institution of Washington, Washington, DC.

Conservation-International, 2007. Biodiversity Hotspots: California Floristic Province.

de Wilde-Duyfjes, B.E.E., 1976. A revision of the genus *Allium* L. (Liliaceae) in Africa. Meded Landbouwhogesch Wageningen 76.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19, 11–15.

Dubouzet, J.G., Shinoda, K., 1999. Relationships among Old and New World Alliums according to ITS DNA sequence analysis. Theoretical Applied Genetics 98, 422–433.

Dubouzet, J.G., Shinoda, K., 1998. Phylogeny of *Allium* L. subgenus *Melanocrommyum* (Webb et Berth.) Rouy based on DNA sequence analysis of the internal transcribed spacer region of nrDNA. Theoretical Applied Genetics 97, 541–549.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Friesen, N., Fritsch, R.M., Blattner, F.R., 2006. Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA ITS sequences. Aliso 22, 372–395.

Hanelt, P., Schultze-Motel, J., Fritsch, R.M., Kruse, J., Maas, H.I., Ohle, H., Pistrick, K., 1992. Infrageneric grouping of *Allium*—the Gatersleben approach. In: Hanelt, P., Hammer, K., Knupffer, H. (Eds.), The Genus *Allium*—Taxonomic Problems and Genetic Resources. IPK, Gatersleben, Germany, pp. 107–123.

Hickman, J.C., 1993. The Jepson Manual: Higher Plants of California. University of California Press, Berkeley, CA.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Klaas, M., 1998. Applications and impact of molecular markers on evolutionary and diversity studies in the genus *Allium*. Plant Breeding 117, 297–308.

Klass, M., Friesen, N., 2002. Molecular markers in Allium. In: Rabinovich, H.D., Currah, L. (Eds.), Allium Crop Science Recent Advances. CABI Publishing, pp. 159–185 (chapter 4).

Kruckeberg, A.R., 1984. California serpentines: Flora, Vegetation, Geology, Soils and Management Problems. University of California Press, Berkeley, CA.

Linne von Berg, G., Samoylov, A., Klaas, M., Hanelt, P., 1996. Chloroplast DNA restriction analysis and the infrageneric grouping of *Allium* (Alliaceae). Plant Systematics and Evolution 200, 253–261.

Maddison, D.R., Maddison, W.P., 2003. MacClade. Sinauer Associates, Inc.

Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis.

McNeal, D.W., 1992. Taxonomy of North American species of *Allium*. In: Hanelt, P., Hammer, K., Knupffer, H. (Eds.), The Genus *Allium*—Taxonomic Problems and Genetic Resources. Institute of Genetics and Crop Plant Research, Gatersleben, Germany, pp. 195–204.

McNeal, D.W., Jacobsen, T.D., 2002. *Allium* L. In: F. o. N.A.E. Committee (ed.), Flora of North America. NYBG Press, New York, USA.

Mes, T.H., Fritsch, R.M., Pollner, S., Bachmann, K., 1999. Evolution of the chloroplast genome and polymorphic ITS regions in *Allium* subgenus *Melanocrommyum*. Genome 42, 237–247.

O'Dell, R.E., Claassen, V.P., 2006. Serpentine and nonserpentine *Achillea millefolium* accessions differ in serpentine substrate tolerance and response to organic and inorganic amendments. Plant and Soil 279, 253–269.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Saghir, A.R.B., Mann, L.K., Ownbey, M., Berg, R.Y., 1966. Composition of volatiles in relation to taxonomy of American Alliums. American Journal of Botany 53, 477–484.

Samoylov, A., Friesen, N., Pollner, S., Hanelt, P., 1999. Use of chloroplast polymorphisms for the phylogenetic study of *Allium* subgenera *Amerallium* and *Bromatorrhiza* (Alliaceae) II. Feddes Repert 110, 103–109.

Samoylov, A., Klaas, M., Hanelt, P., 1995. Use of chloroplast polymorphisms for the phylogenetic study of subgenera *Amerallium* and *Bromatorrhiza* (genus *Allium*). Feddes Repert 106, 161–167.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Systematic Biology 49, 369–381.

Stern, W.T., 1992. How many species of *Allium* are known? Kew Magazine 9, 180–182.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Traub, H.P., 1968. The subgenera, sections and subsection of *Allium* L. Plant Life 24, 147–163.

Traub, H.P., 1972. Genus *Allium* L.—subgenera, sections and subsections. Plant Life 28, 132–137.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. PCR Protocols: A Guide to Methods and Applications.. Academic Press, Inc.

Wright, J.W., Stanton, M.L., Scherson, R., 2006. Local adaptation to serpentine and non-serpentine soils in *Collinsia sparsiflora*. Evolutionary Ecology Research 8, 1–21.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Please cite this article in press as: Nguyen, N.H. et al., A molecular phylogeny of the wild onions (*Allium*; Alliaceae) ..., Mol. Phylogenet. Evol. (2008), doi:10.1016/j.ympev.2007.12.006