Molecular Phylogenetics and Evolution xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution

Ann Willyard [a,*,1], Lisa E. Wallace [b,1], Warren L. Wagner [c], Stephen G. Weller [d], Ann K. Sakai [d], Molly Nepokroeff [a]

[a] University of South Dakota, Department of Biology, 414 E. Clark St., Vermillion, SD 57069, USA
[b] Mississippi State University, Department of Biological Sciences, P.O. Box GY, Mississippi State, MS 39762, USA
[c] National Museum of Natural History, Smithsonian Institution, Botany, P.O. Box 37012, MRC 166, Washington, DC 20013, USA
[d] University of California Irvine, Department of Ecology & Evolutionary Biology, 321 Steinhaus Hall, Irvine, CA 92697, USA

ARTICLE INFO

Article history:
Received 1 August 2010
Revised 4 April 2011
Accepted 5 April 2011
Available online xxxx

Keywords:
Hawai'i
Species trees
Reticulate evolution
Schiedea
Caryophyllaceae
Low-copy nuclear loci

ABSTRACT

Schiedea (Caryophyllaceae) is a monophyletic genus of 34 species, all endemic to the Hawaiian Islands, that arose from a single colonization, providing one of the best examples of adaptive radiation in Hawai'i. Species utilize a range of habitats and exhibit a variety of growth forms and transitions in breeding systems from hermaphroditism toward dimorphism or autogamy. Our study included the most thorough sampling to date: 2–5 individuals per species and 4 independent genetic partitions: eight plastid and three low-copy nuclear loci (9217 bps), allowing a three-locus BEST species tree. Despite incomplete resolution at the tips, our results support monophyly for each extant species. Gene trees revealed several clear cases of cytonuclear incongruence, likely created by interspecific introgression. Conflict occurs at the divergence of section *Alphaschiedea* as well as at the tips. Ages inferred from a BEAST analysis allow an original colonization onto either Nihoa or Kauaì and inform some aspects of inter-island migrations. We suggest that several hard polytomies on the species tree are biologically realistic, signifying either nearly simultaneous speciation or historical introgressive hybridization. Based on inferred node ages that exceed expected coalescent times, we propose that undetected nuclear introgression may play a larger role than incomplete lineage sorting in sections *Schiedea* and *Mononeura*.

1. Introduction

1.1. Species tree construction

Adaptive radiations are a major challenge to molecular systematics because individual gene lineages may be so recent that they fail to coalesce before the time of species divergence (Edwards et al., 2007). Among recently diverged species, genealogies inferred from independent regions of the genome are likely to disagree due to the differential sorting of ancestral polymorphism into daughter lineages such that each gene tree may differ from the species tree (Degnan and Rosenberg, 2006; Funk and Omland, 2003; Goodman et al., 1979; Maddison, 1997). To construct a species tree where coalescence cannot safely be assumed, multiple independent loci are vital to overcome the stochastic nature of lineage sorting (Edwards et al., 2007; Kubatko and Degnan, 2007; Maddison and Knowles, 2006; Rokas et al., 2003). Multiple accessions per species may also strengthen inferences if young species groups still share polyphyletic alleles (Edwards and Beerli, 2000; Maddison and Knowles, 2006). Although incomplete lineage sorting is a well-recognized feature of species-level phylogenies, methods to infer species trees from disparate gene trees are in their infancy (e.g. Carstens and Knowles, 2007; Edwards, 2009; Huang and Knowles, 2009; Liu et al., 2008; Maddison, 1997). Despite simulations that describe inconsistent results from concatenated data in groups with rapid and/or recent radiations (Kubatko and Degnan, 2007; Liu and Edwards, 2009), 'single-tree' techniques are still widely used. Species tree approaches that simultaneously consider each gene tree have the potential to provide a useful integration of population genetics and phylogenetics, particularly where

Abbreviations: π, nucleotide diversity; AIC, Akaike information criterion; BS, bootstrap; BEAST, Bayesian evolutionary analysis by sampling trees; BEST, Bayesian estimation of species trees; CI, consistency index; GSI, genealogical sorting index; Ma, million years ago; MCMC, Markov chain Monte Carlo simulations; PI, parsimony-informative; PP, posterior probability; RI, retention index; TL, tree length; TMRCA, time to most recent common ancestor.

* Corresponding author. Address: Hendrix College, Biology Department, 1600 Washington Ave., Conway, AR 72032, USA. Fax: +1 501 450 4547.

E-mail addresses: willyard@hendrix.edu (A. Willyard), lisawallace@biology.msstate.edu (L.E. Wallace), wagnerw@si.edu (W.L. Wagner), sgweller@uci.edu (S.G. Weller), aksakai@uci.edu (A.K. Sakai), Molly.Nepokroeff@USD.edu (M. Nepokroeff).

[1] These authors contributed equally to this work.

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.04.001

conditions of high gene tree discordance due to incomplete lineage sorting are to be expected, such as those encountered in recent, rapid radiations (Edwards et al., 2007). However, their implementation can be challenging, and few empirical studies have used species-tree methods for large numbers of samples (Linnen and Farrell, 2008).

1.2. Reticulate evolution

Reticulate evolution arises if genetic information is acquired from a divergent taxon and a remnant of the 'foreign' genome persists in a lineage. Several mechanisms for such gene transfers are known from eukaryotes (Doolittle and Bapteste, 2007), and interspecific hybridization that leads to the persistence of introgressed genomic regions (introgression) appears to be a common phenomenon in the divergence of plant species (Grant, 1981; Seehausen, 2004; Stebbins, 1950). Species barriers are particularly porous to the introgression of maternally inherited organelle genes (Chan and Levin, 2005). Reports of introgressed mitochondrial genes are frequent (Chan and Levin, 2005; Shaw, 2002), and introgressed plastid genes appear to be nearly ubiquitous across the plant kingdom (Rieseberg and Soltis, 1991). Factors that make organellar introgression much more likely than nuclear include maternal inheritance (Chan and Levin, 2005; Rieseberg et al., 1996), lack of linkage to nuclear genes under selection (Funk and Omland, 2003) and smaller effective population size due to clonal inheritance (Petit et al., 1993). Despite a recognition that reticulate evolution appears to be a common phenomenon (especially, but not exclusively, in plant species), the complexity of the problem has so far prevented chloroplast capture from being included in models used for phylogenetic algorithms that infer species trees (Edwards et al., 2007). While the exclusive use of organellar genes to construct species trees is misleading in the case of chloroplast capture, their uniparental inheritance can be helpful in conjunction with nuclear genes to infer the direction, as well as the relative timing, of introgression, (e.g. Bossu and Near, 2009; Dunbar-Co et al., 2008; Howarth and Baum, 2005; Lindqvist et al., 2003; Willyard et al., 2009).

1.3. Schiedea overview

Monophyly in *Schiedea* (Caryophyllaceae), a genus of herbs, vines, and small shrubs endemic to the Hawaiian Islands, suggests that all 34 known species (32 of them extant) arose *in situ* following a single colonization of the island chain (Wagner et al., 2005). *Schiedea*'s closest sister genera (*Honckenya* and *Wilhelmsia*) are circumboreal (Wagner et al., 2005) with the nearest populations growing some 4000 km from the Hawaiian archipelago. Species of *Schiedea* are unique among the Caryophyllaceae in possessing distinctive floral nectaries. Developmental studies have shown that these nectaries, which may have co-evolved with honeycreeper and/or honeyeater bird pollinators in sections *Nothoschiedea* and *Alsinidendron* (Weller et al., 1998), and reduced structures in wind-pollinated *Schiedea*, represent homologous organs across the genus (Wagner et al., 2005). The presence of a woody habit in some species also delineates *Schiedea* from other members of the subfamily Alsinoideae (Wagner et al., 1999), and is one the numerous examples of woodiness arising in island flora from herbaceous progenitors (Carlquist, 1974). *Schiedea* represents the fourth most species-rich radiation of plants in the Hawai'ian Islands, and the third from a single colonization of these islands (Wagner et al., 1999). Fifteen of the 32 extant species are endangered, giving *Schiedea* the undesirable distinction of having the "highest proportion of endangered taxa for any species-rich lineage in the Hawaiian Islands" (Wagner et al., 2005). Most *Schiedea* species are single-island endemics. When Lana'i, Moloka'i, and Maui are considered as a single land mass (Maui Nui) due to geologically recent and prolonged above-water connections (Wagner et al., 1999), only four *Schiedea* species have populations that occupy multiple islands. This high level of single-island endemism in a species-rich genus provides the opportunity to compare reproductive isolation achieved via inter-island migrations with allopatric speciation created by habitat shifts within an island, an objective possible only with a resolved phylogeny for the lineage (Sakai et al., 2006; Weller and Sakai, 1999). A resolved phylogeny would also be useful for understanding the number of breeding systems transitions that have occurred in this lineage, and possibly the ecological factors that have driven these changes.

1.4. Timing of Schiedea colonization and radiation

The age of all *Schiedea* species may be constrained by the age of the current high Hawai'ian Islands, i.e., within the last 5–7 million years (Ma; Clague, 1996). The recent nature of the *Schiedea* radiation is supported by low interspecific nucleotide sequence divergence in nuclear ribosomal (nrDNA) internal and external transcribed spacers (*ITS* and *ETS*; Wagner et al., 2005). However, intraspecific variation has not been tested across the genus, as previous genus-wide molecular systematic studies have used single-exemplar sampling. Based on the geographic distribution of most apparently early-branching *Schiedea* lineages (Wagner et al., 2005), the original immigration may have been onto the oldest current high island in the chain (Kauaì; ca. 4.7 Ma; Clague, 1996), placing a lower boundary for the origin of *Schiedea* within that time frame. Alternatively, the presence of a single *Schiedea* species (*S. verticillata*) on the eroded island of Nihoa could represent the persistence of the lineage from one of the older, leeward, islands while it was higher. The timing and spacing of volcanic island formation in this chain may have provided continuous habitat as long ago as 30 or 23 Ma, dating to times when the atolls of Kure or Lisianski, respectively, were high islands (Clague, 1996). 'Island-hopping' to a new island before all of the suitable habitat on older islands was degraded has been proposed for lineages of some organisms on the Hawaiian archipelago (Givnish et al., 2009, 1995; Jordan et al., 2003; Russo et al., 1995; Schneider et al., 2005), while other colonizations have been described as being more recent and thus their radiations more rapid (Fleischer et al., 1998; Havran et al., 2009; Price and Clague, 2002). A resolved *Schiedea* species tree would help calibrate the age of the genus and thus improve estimations for the rate of speciation in the lineage.

1.5. Evolution of breeding systems

The genus *Schiedea* is of additional interest as an outstanding system in which to study diversification of sexual dimorphism in plants. Dioecy arose within the genus (Wagner et al., 2005), and a wide range of breeding systems are currently represented: 14 species are moderately to highly outcrossing hermaphrodites, eight are facultatively autogamous or cleistogamous, and 10 are dimorphic: gynodioecious, subdioecious or dioecious (Wagner et al., 2005; Weller et al., 1998). *Schiedea* has been used to test some of the factors that promote dioecy, e.g. inbreeding depression (Sakai et al., 1989, 1997), sex allocation (Sakai et al., 2008; Weller and Sakai, 2005), wind pollination (Golonka et al., 2005; Weller et al., 1998, 2006, 2007), and environmental conditions (Culley et al., 2006; Sakai and Weller, 1991). However, the number of shifts to dimorphic breeding systems within *Schiedea* is unknown, as it depends on phylogenetic resolution among the 12 very closely related species in section *Schiedea*. Up to six independent transitions to dioecy have been suggested within this section (Norman et al., 1997; Weller et al., 1990, 1995), and previous phylogenies have not provided any evidence that

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

the three hermaphroditic species in section *Schiedea* retain that feature as the plesiomorphic state. At least one reversal toward hermaphroditism may have occurred: from gynodioecy to hermaphroditism in *S. lydgatei* (Soltis et al., 1996). In *Schiedea*, male sterility (a critical component of dioecy) is under the control of a single, recessive nuclear gene (Weller and Sakai, 1991). This suggests that the allele conferring male sterility might be passed among species if interspecific hybridization leads to the introgression of some nuclear genes between different species. A better understanding of the phylogeny within section *Schiedea* will allow an investigation of whether male sterility arose as a single transition followed by retention of the allele in the lineage with dimorphic species, an evolutionary scenario suggested by apparent identity of the male sterility gene across species (Weller et al., 2001). Alternatively, the male sterility allele may have been transferred across species boundaries by hybridization. A resolved phylogeny would also help determine if any lineages within section *Schiedea* truly represent reversals toward hermaphroditism.

1.6. Introgression among Schiedea

The role of interspecific gene flow in the diversification of plant and animal adaptive radiations, (reviewed by Seehausen (2004)) has been recently explored in a number of plant and animal groups (e.g. (Caraway et al., 2001; Frajman et al., 2009; Howarth and Baum, 2005; Leache, 2009; Pirie et al., 2009; Taylor and McPhail, 2008)). We suspected that introgression might have occurred within *Schiedea*. Based on morphological intermediates, natural hybridization has been reported for five pairs of *Schiedea* species: extensive numbers of hybrid plants for *S. ligustrina* × *S. mannii*, *S. lydgatei* × *S. sarmentosa*, and *S. menziesii* × *S. salicaria*; minimal numbers for *S. hookeri* × *S. mannii* and *S. hookeri* × *S. pentandra* (Wagner et al., 2005; Weller et al., 2001). Introgression of plastid genes was inferred from molecular evidence between *S. kealiae* × *S. ligustrina* (Soltis et al., 1996). In contrast to the apparently minimal occurrence of natural hybridization, vigorous F_1 progeny with fertile pollen can be created from artificial crosses between many *Schiedea* species (Weller et al., 2001). This suggests that allopatry provides reproductive isolation in *Schiedea*, as most species are currently limited to specialized habitat within a fairly small range that rarely coincides with the range and/or the habitat of congeneric species (Weller et al., 2001). Opportunities for sympatric or nearly sympatric pairs of *Schiedea* to hybridize are further reduced by the tendency toward autogamy in eight species. For example, no morphological intermediates have been detected between *S. nuttallii* and the sympatric but facultatively autogamous *S. obovata*, between *S. pentandra* and the sympatric but cleistogamous *S. trinervis*, or between *S. stellarioides* and the sympatric but facultatively autogamous *S. viscosa*. The extent of hybridization between other sympatric pairs (e.g. *S. kauaiensis* × *S. membranacea*, *S. spergulina* × *S. stellarioides*) is an open question, as no morphological intermediates have been reported.

1.7. Experiment overview

Species-tree methods are computationally difficult with a large number of samples. Our data set consisted of 91 accessions (yielding 182 nuclear alleles) to adequately sample all 32 extant species of *Schiedea* across their geographic ranges, and to address morphological variation. For 91 tips, the number of possible tree rearrangements approaches 3×10^{195} (Felsenstein, 2004), a vast number for any phylogenetic method, let alone one that takes individual gene tree topologies into account. In this paper, we use our multi-locus data set that represents both plastid and nuclear genomes to compare inferences between a species tree approach and a traditional concatenation method. We also use a 'total-evidence' analysis that concatenates nuclear, plastid, and morphological data. In overview, our process was to: (1) create four independent gene trees – one for plastid loci and one for each of three independent nuclear loci; (2) create a 'Bayesian Estimation of Species Trees' (BEST; Liu, 2008) using three nuclear loci; (3) concatenate three nuclear loci and create a nuclear single-tree species estimate; (4) evaluate species monophyly and identify accessions with conflicting placement on plastid and nuclear gene trees; (5) exclude the plastid information for species with cytonuclear conflict to create a 'compromise' data set; (6) create a second BEST species tree that combines nuclear and plastid data using this compromise data set; (7) test the total-evidence approach by analyzing a 32-allele data set (applying the same compromise for cytonuclear conflict) that collapses the multi-accession data set to one record per species and incorporates previously published nrDNA sequences and morphology. The total-evidence process retains much of the valuable data from plastid loci and incorporates additional information for each species, while attempting to avoid major violations of the assumption of hierarchical evolution. We compare species tree results from these very different approaches to test the utility of species-tree vs. concatenated methods and the reliability of identifying and suppressing cytonuclear conflict in this species-rich young lineage.

We address six main questions: (1) Are named taxonomic species of *Schiedea* monophyletic? (2) How do traditional sectional relationships compare to relationships resolved on our species trees? (3) Given multiple loci and the apparently rapid and recent nature of the *Schiedea* lineage, what is the relative performance of the species-tree vs. the concatenated method for inferring species relationships in *Schiedea*? (4) At what evolutionary depth and timeframe do incomplete lineage sorting and introgression result in statistically inconsistent species trees within the *Schiedea* lineage? (5) Can we improve species tree inference by identifying and excluding introgressed alleles and by incorporating additional data for each species? (6) What is the age of the genus and of the most recent crown radiation?

2. Materials and methods

2.1. Plant materials

Ninety-one individuals were sampled from 51 populations, representing all 32 extant *Schiedea* species with two or three accessions per species except that five accessions were included for *S. globosa* in order to sample populations from the islands of Hawai'i, Maui, and Oàhu (Table 1). Herbarium abbreviations follow the online database *Index Herbariorum* (Thiers, 2010). Nucleotide sequences for *S. globosa*, *S. haleakalensis*, *S. hookeri*, and *S. mannii* were included from a previous study: GenBank accessions FJ496359, FJ496372, FJ496374, FJ496387, FJ496413, FJ496416- FJ496425, FJ496430, FJ496455, FJ496458, FJ496459, FJ496481, FJ496526, FJ496527, FJ496531- FJ496541, FJ496545, FJ496546, FJ496562, FJ496563, FJ496566, FJ496567, FJ496587, FJ496588, FJ496632, FJ496633, and FJ496638- FJ496647 (Wallace et al., 2009). Plants were grown in the UC-Irvine greenhouses from wild-collected cuttings or seed. Harvested leaves were stored in silica or frozen until being processed. Five outgroups were selected from subfamily Alsinoideae (Caryophyllaceae; Table 1).

2.2. DNA isolation, amplification, and sequencing

DNA was isolated with the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). Eleven loci were PCR-amplified and sequenced. These included portions of three low-copy nuclear genes: chloroplast-expressed glutamine synthetase (*ncpGS*), nitrate reductase

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

Table 1
Collection localities, collectors, and vouchers for 91 individual plants for 32 *Schiedea* species and 8 individual plants for five outgroup species.

Species	Number of individuals	Geographic location and voucher
adamantis St. John	3	Oáhu: Diamond Head; cult. Weller & Sakai 847 (BISH)
apokremnos St. John	3	Kaua'i: Napali cliffs; Flynn et al. 2219 (PTBG); cult. Weller & Sakai 865 (BISH)
attenuata W.L. Wagner, Weller and Sakai	3	Kaua'i: Kalalau Valley; Wood 1394 (US)
diffusa A. Gray	1	Hawai'i: Ola'a Tract; Perlman et al. 14780 (PTBG); cult. Wagner & Shannon 6796 (BISH)
	2	E. Maui: Halehaku Lake; Hughes s.n.; cult. Weller & Sakai 910 (BISH)
globosa H. Mann	1	Hawai'i: Paokalani Island; Wood 11221 (PTBG)
	1	Oáhu: Makapu'u Beach; cult. Weller & Sakai 844 (BISH)
	2	Oáhu: Koko Head; cult. Weller & Sakai 906 (BISH)
	1	W. Maui: Pohakupule; cult. Weller & Sakai 951 (US)
haleakalensis Degener and Sherff	2	E. Maui: Haleakala NP; cult. Weller & Sakai 851 (BISH)
hawaiiensis Hillebrand	3	Hawai'i: Pahakuloa Training Area; PTA personnel s.n.; cult. Weller & Sakai 932 (BISH)
helleri Sherff	1	Kaua'i: Mohihi-Wai'alae trail; Perlman & Wood 13572; cult. Wagner & Shannon 6809 (BISH)
	2	Kaua'i: Nawaimaka Valley; Perlman 14656 (PTBG)
hookeri A. Gray	1	Oáhu: between Wai'anae Kai & Makaha valleys; cult. Weller & Sakai 794 (BISH)
	1	Oáhu: Kaluakauila gulch; Wagner s.n.; cult. Weller & Sakai 862 (US)
	1	Oáhu: between Makua & Wai'anae valleys; Obata s.n.; cult. Weller & Sakai 866 (US)
jacobii W.L.Wagner, Weller and Medeiros	2	E. Maui: between Hanawi stream & Kuhiwa drainage; cult. Perlman et al. 14807 (BISH)
kaalae Wawra	1	Oáhu: near Pu'umaialau; Takeuchi 3587 (BISH); cult. Weller & Sakai s.n. (US)
	2	Oáhu: Kalua'a gulch; cult. Weller & Sakai 892 (US)
kauaiensis St. John	2	Kaua'i: Mahanaloa Valley; Perlman & Obata 12074 (BISH); cult. Weller & Sakai s.n. (US)
	1	Kaua'i: Mahanaloa Valley; Wood 7430 (PTBG)
kealiae Caum and Hosaka	3	Oáhu: Kealia trail; cult. Weller & Sakai 791 (BISH)
laui W.L. Wagner and Weller	2	Moloka'i: Waikolu drainage; Lau & Loo 3951 (US)
ligustrina Chamisso and Schlechtendal	1	Oáhu: NE of Palikea Peak; Obata & Perlman s.n.; cult. Weller & Sakai 846 (US)
	2	Oáhu: Lualualei Magazine; cult. Weller & Sakai 873 (US)
lychnoides Hillebrand	3	Kaua'i: Alaka'i Swamp; Flynn s.n.; cult. Weller & Sakai 867 (US)
lydgatei Hillebrand	2	Moloka'i: Pu'ukolekole; Wagner et al. 6682; cult. Wagner & Shannon 6856 (BISH)
	1	Moloka'i: Kamiloloa trail; cult. Weller & Sakai 870 (BISH)
mannii St. John	1	Oáhu: between Wai'anae Kai & Makaha valleys; cult. Weller & Sakai 793 (BISH)
	1	Oáhu: Pu'u kawiwi; cult. Weller & Sakai 901; cult. Wagner & Shannon 6813 (BISH)
membranacea St. John	2	Kaua'i: junction of Mahanaloa & Ku'ia valleys; cult. Weller & Sakai 864 (BISH)
menziesii Hooker	2	W. Maui: Lihau Ridge; cult. Weller & Sakai 849 (BISH)
	1	W. Maui: Hahakea gulch; cult. Weller & Sakai 949 (US)
nuttallii Hooker	1	Oáhu: Makua gulch; Perlman 11046; cult. Weller & Sakai 861 (BISH)
	2	Oáhu: Keawapilau gulch; cult. Weller & Sakai 903 (BISH)
obovata (Sherff) W.L. Wagner and Weller	3	Oáhu: Makua gulch; cult. Weller & Sakai 868 (US)
pentandra W.L. Wagner and E. Harris	2	Oáhu: Palikea trail; cult. Weller & Sakai 860 (US)
perlmanii W.L. Wagner and Weller	3	Kaua'i: Ha'upu; Perlman et al. 12917 (PTBG); cult. Wagner & Shannon 6795 (BISH)
pubescens Hillebrand	1	Moloka'i: Waianui gulch; Perlman 14682 (BISH);
	1	Moloka'i: Waianui gulch; Wood & Perlman 4088 (US)
salicaria Hillebrand	2	W. Maui: Pu'uhona; cult. Weller & Sakai 842 (BISH)
	1	W. Maui: Ka'onohua gulch; cult. Weller & Sakai 853 (US)
sarmentosa Degener and Sherff	3	Moloka'i: Makolelau gulch; cult. Weller & Sakai 896 (US)
spergulina A. Gray	3	Kaua'i: Waimea Canyon; cult. Weller & Sakai 863 (BISH)
stellarioides H. Mann	2	Kaua'i: between Wai'alae & Nawaimaka valleys; Perlman & Wood 14651; cult. Wagner & Shannon 6855 (BISH)
	1	Kaua'i: Wai'alae Valley; Perlman 14765; cult. Wagner & Shannon 6854 (BISH)
trinervis (H. Mann) Pax and K. Hoffmann	1	Oáhu: between Kalena & Mt. Ka'ala; Perlman 5448 (BISH); cult. Weller & Sakai s.n. (US)
	2	Oáhu: Mt. Ka'ala; Weller & Sakai 908; cult. Wagner & Shannon 6853 (BISH)
verticillata F. Brown	1	Nihoa: Devil's Slide & Dog's Head; Conant et al. s.n.; cult. Wagner & Shannon 6819 (BISH)
	2	Nihoa: Devil's Slide & Dog's Head; Conant et al. s.n.; cult. Weller & Sakai 880
viscosa H. Mann	3	Kaua'i: Mohihi-Wai'alae trail; Flynn 5031 (PTBG); cult. Wagner & Shannon 6810 (BISH)
Honckenya peploides (L) Ehrh.	3	Oregon, USA: W.L. Wagner 6915 (US); Alaska, USA: Weller and Sakai 974 (US); Baffin Island, Canada: Gillespie et al. 6784 (US)
Wilhelmsia physodes Fisch ex DC	2	Alaska, USA: C.L. Parker 8422 (US); Alaska, USA: Grant 3739 (ALA)
Minuartia moehringioides (DC) Mattf.	1	Durango, Mexico: González 6259 (CIIDIR)
Scleranthus annuus L.	1	Washington DC, USA: W.L. Wagner 6862 (US)
Scleranthus biflorus Hook. f.	1	Wellington, New Zealand: Smisson WELTU 19660 (WELTU)

(*NIA*), and the fourth intron of phosphoenolpyruvate carboxylase (*PepC*) and eight noncoding regions of the chloroplast genome, selected for their relatively high variability in other angiosperms: six intergenic regions (*petN–psbM*, *psbE–petL*, *psbM–trnD*, *trnD–trnT*, and *trnQ–rpS16*); a portion of the *tRNA–Lys* (*trnK*) intron with primers in 5′*trnK* and 5′*matK;* a portion of the *trnK* intron with primers in 3′*matK* and 3′*trnK;* and the *RPS16* intron. PCR primers were previously published (Demesure et al., 1995; Emshwiller and Doyle, 1999; Gaskin and Schaal, 2002; Howarth and Baum, 2002; Johnson and Soltis, 1994; Lee and Wen, 2004; Popp et al., 2005; Popp and Oxelman, 2001; Shaw et al., 2005, 2007) or designed for this study (Supplementary Table S1).

Each 25 μl PCR reaction contained 1 μl of template DNA, 0.2 mM each primer, 0.5 mM each dNTP, 2.5 mM $MgCl_2$, 1 μl BSA (10X), and 1.0 U of GoTaq® Flexi DNA polymerase in supplied colorless buffer (Promega, Madison, WI, USA). For *PepC*, reaction sizes were 50 μl, with the same concentrations except that 2 μL BSA (10X), 1.0 U DNA polymerase, and 5.0 μL template DNA were used. For the nuclear loci, the thermal cycler program preheated at 95 °C for 2 min, and then ran 35 cycles of denaturing for 1 min at 95 °C, annealing at 50 °C with a ramp of 0.1 °C per second to 65 °C for 1 min, and extending for 1 min at 65 °C. For the plastid loci, the thermal cycler program preheated at 95 °C for 5 min, and then ran 35 cycles of denaturing for 1 min at 95 °C, annealing at 50 °C

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

for 1 min, extending for 2 min at 72 °C, and a final extension for 7 min at 72 °C.

A sample of each PCR product and the negative controls were electrophoresed in TBE agarose gels to check for amplification of a band of the expected size and for the absence of contamination. Plastid PCR products were cleaned with the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). For nuclear loci, bands cut from gels were purified with the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). Cleaned DNA was sequenced from the forward and reverse primers with the dideoxy chain termination method using Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and Better Buffer (The Gel Company, San Francisco, CA, USA). For two longer amplicons (*psbE–petL* and *trnD–trnT*), additional sequencing was performed from internal primers (Supplementary Table S1). Reactions were cleaned with an ethanol/sodium acetate precipitation, and samples were capillary electrophoresed in an Avant-3100 Genetic Analyzer following manufacturer's protocols (Applied Biosystems, Foster City, CA, USA).

2.3. Data preparation

Raw sequences were manually edited and a contig of each allele was created in Sequencher vers. 4.1 (Gene Codes, Ann Arbor, MI, USA). For individuals heterozygous for *ncpGS*, *NIA*, or *PepC*, we inferred the identities of the two alleles using haplotype subtraction (Clark, 1990). For those individuals whose allelic identities could not be determined by this method, we cloned that locus with the *pGEM*-T Easy Vector System II (Promega Madison, WI, USA) and sequenced individual clones as described. The two-allele data set included all 182 sequences, even if identical. The same arbitrary selection of one allele per accession (91 samples) was used for all one-allele analyses. Sequences were manually aligned using Se–Al vers. 2.0 (Rambaut, 2010b) to minimize the number of inferred indels. Nucleotide diversity (π; Nei, 1987) was calculated across the one-allele data set as the average number of nucleotide differences per site between two sequences, excluding sites with missing data, using the software DnaSP vers. 4.5.0 (Rozas et al., 2003). For *NIA*, we excluded one missing allele from *S. apokremnos* before calculating π.

Recombinant sequences can distort phylogenetic inferences. We tested for evidence of recombination in the *ncpGS*, *NIA*, and *PepC* nuclear alignments by applying two statistical tests, the RDP method (Martin and Rybicki, 2000) and the MaxChi method implemented in the software RDP vers. 2.0 (Martin et al., 2005). For the RDP test, we analyzed the sequences at window sizes of 5, 10, 50, and 100, using internal references. The MaxChi test was run by considering triplets of sequences and all sequences simultaneously with gaps removed and a variable window size of 0.013. The significance of χ^2 peaks was tested using a permutation test of 1000 iterations.

Because we have a data set with multiple accessions per species, we are able to examine the monophyly of each named taxonomic species. The genealogical sorting index (*gsi*) is a standardized measure of exclusive ancestry based on a tree topology. The index accounts for polytomies that do not conflict with shared ancestry and is normalized to account for differences in sample size among clades (Cummings et al., 2008). Index values range between $gsi = 1$ (complete monophyly) and $gsi = 0$ (no shared ancestry). We calculated the *gsi* for each species for each of our four Bayesian majority-rule gene trees (plastid, *ncpGS*, *NIA*, and *PepC*), selecting the one-allele gene tree for nuclear loci, using a web-based application (Cummings et al., 2010). For each statistic, 1000 permutations were used to assess significance at the $P = 0.05$ level. An ensemble *gsi* statistic was calculated using all four of our gene trees to provide an overall estimate of each species' monophyly. This single statistic accounts for the uncertainty on one or more trees by weighting the *gsi* value from each tree with its probability. We considered the *gsi* statistic when deciding which species had paraphyletic, and therefore potentially introgressed, plastid genes that should be excluded in the compromise species tree.

After preliminary investigations, we selected a single accession, *Honckenya peploides* (W.L. Wagner 6915), as the outgroup for all analyses, because the current facility in BEST only allows for a single outgroup. For all phylogenetic analyses, alignment gaps were treated as missing data. We used SeqState vers. 1.4 (Müller, 2005) to recode gaps between ingroup alleles in our alignments as presence/absence characters using the simple indel-coding method (Simmons and Ochoterena, 2000). Indel-coding characters were added to the matrices used for parsimony and Bayesian analyses, with these characters coded as missing data for the outgroup in order to minimize homoplasy.

2.4. Phylogenetic analyses

Model parameters are summarized in Table 2. We used similar methods to infer individual gene trees and for concatenated species trees. Parsimony searches were conducted in PAUP* vers. 4.0b10 (Swofford, 2002) using a heuristic search with 10 independent repetitions of random sequence addition and the tree bisection reconnection method of branch swapping. For some analyses (Table 2), the maximum number of trees held was limited to 100,000 in order to reduce computation time. Nodal support was estimated using 1000 bootstrap pseudoreplicates. Based on the findings of DeBry and Olmstead (2000), one tree was held during bootstrapping for some analyses (Table 2).

A nucleotide substitution model was selected with the Akaike information criterion (AIC; Akaike, 1974) from among 88 possible models using jModeltest vers. 2.0 (Posada, 2008). The best-fit model for the concatenated eight-locus plastid alignment and for *ITS*/*ETS* is GTR + G. Using the two-allele data set, the best-fit model is GTR + G for *PepC*, TMP3uf + G for *ncpGS*, and GTR + I + G for *NIA*. Thus, our nucleotide partitions in Bayesian analyses allowed six substitution types for every partition. Among-site variation was modeled with a gamma distribution for plastid, *ITS*/*ETS*, *PepC*, and *ncpGS* partitions, while the *NIA* partition allowed a proportion of sites to be invariable with the remainder gamma-distributed.

Bayesian analyses to infer gene trees and concatenated species trees were performed by allowing rates to vary by partition. For data included in a particular analysis (Table 2), we defined up to seven partitions: (1) the eight plastid loci as a single partition; (2–4) a partition for each low-copy nuclear locus individually; (5) a partition for all indel-coding characters; (6) a single partition for nrDNA *ITS* and *ETS* sequences; and (7) a partition for morphological characters. Bayesian searches applied the optimum substitution model to each nucleotide partition and an equal-rate binary model to partitions for morphological and indel characters. For indel characters, the coding parameter was set to indicate that only variable characters had the possibility of being sampled. For each analysis, we performed two simultaneous runs using a multi-processor compilation of MrBayes vers. 3.1.2 (Ronquist and Huelsenbeck, 2003). The number of chains of Markov chain Monte Carlo simulations (MCMC), number of generations, and temperature parameters for chain swapping are given in Table 2. All runs were sampled at the frequency needed to save 20,000 trees per run. We evaluated convergence and chose the number of samples to discard as burn-in based on stationarity of a plot of the generation versus log-likelihood for each run. We also compared plots of tree distances and split frequencies between two runs using the 'Comparetree' function in MrBayes. One majority-rule consensus tree was built for each data set by combining trees generated by

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

Table 2
Summary of data, parameters, and results for gene trees and species trees; BS, bootstrap; CI, consistency index; MCMC, Metropolis–coupled Monte Carlo simulations PI, parsimony-informative; RI, retention index; TL, tree length.

Data set	Ingroup alleles	Aligned nucleotides (PI)	Indel codes (PI)	Morphological characters (PI)	Total (PI)	Methods and results summary
Eight plastid loci[a]	91	6793 (261)	117 (68)	0	6910 (329)	Bayesian: GTR + G, 4 chains MCMC, 3 million generations, temperature = 0.25; Fig. 1 Parsimony: maxtrees = 100,000; BS maxtrees = 1; 100,000 trees of TL = 768; CI = 0.8320; RI = 0.9518 BEAST[b]: prior ages at 7.3, 4.7, and 3.7 (Fig. 6) or at 4.7 and 3.7 (Fig. 7)
ncpGS	182	845 (321)	56 (36)	0	901 (357)	Bayesian: TPM3uf + G, six chains MCMC, 20 million generations, temperature = 0.30; Supplementary Fig. S1
	91[c]	845 (85)	23 (11)	0	868 (96)	Bayesian: TPM3uf + G, six chains MCMC, 20 million generations, temperature = 0.30 Parsimony: maxtrees = 100,000; BS maxtrees = 1;100,000 trees of TL = 473; CI = 0.8055; RI = 0.8504
NIA	180	964 (295)	124 (72)	0	1088 (367)	Bayesian: GTR + I + G, 6 chains MCMC, 20 million generations, temperature = 0.30; Supplementary Fig. S2
	90[c,d]	964 (214)	101 (48)	0	1065 (262)	Bayesian: GTR + I + G, 6 chains MCMC, 20 million generations, temperature = 0.30 Parsimony: maxtrees = 100,000; BS maxtrees = 1; 100,000 trees of TL = 1015; CI = 0.6148; RI = 0.8363
PepC	182	615 (200)	36 (35)	0	651 (235)	Bayesian: GTR + G, six chains MCMC, 20 million generations, temperature = 0.30; Supplementary Fig. S3
	91[c]	615 (64)	25 (11)	0	640 (75)	Bayesian: GTR + G, six chains MCMC, 20 million generations, temperature = 0.30 Parsimony: maxtrees = 100,000; BS maxtrees = 1; 100,000 trees of TL = 297; CI = 0.8552; RI = 0.9042
Three nuclear loci	91	2424 (363)	0	0	2424 (363)	Best: nuclear species tree from three gene trees (*ncpGS*; *NIA*; *PepC*); Fig. 2 Bayesian concatenated: 3 partitions (*ncpGS*; *NIA; PepC*), 6 chains MCMC, 20 million generations, temperature = 0.30; Fig. 3
Total-evidence[b]	32[e]	10,506 (765)	128 (67)	61 (56)	10,694 (888)	Parsimony tree: one best tree of TL = 2309; CI = 0.6003; RI = 0.7142; Fig. 4a Bayesian concatenated: seven partitions (eight plastid loci; *ncpGS*; *NIA; PepC*; *ITS/ETS*; indel codes; morphology), six chains MCMC, 20 million generations, temperature = 0.30; Fig. 4b
			0	0	10,506 (765)	Maximum likelihood tree: five partitions (eight plastid loci; *ncpGS*; *NIA; PepC*; *ITS/ETS*); Fig. 4c
Eight plastid + 3 nuclear loci[b]	91[c]	9217 (624)	0	0	9217 (624)	Best: nuclear-plastid species tree from four partitions (eight plastid loci; *ncpGS*; *NIA*; *PepC*); Fig. 4d

[a] see Supplementary Table S1 for plastid loci.
[b] Using the compromise adjustment for cytonuclear conflict (see Section 3.3).
[c] The same set of one randomly-selected nuclear allele per accession was used in every analysis.
[d] One accession of *S. apokremnos* is missing from *NIA* data set.
[e] One consensus record per species.

two runs, discarding trees prior to stationarity. Branch lengths were estimated by averaging across all retained trees. Maximum likelihood analyses for the concatenated 32-allele data set were run in RAxML vers. 7.0.4 (Stamatakis, 2006) using a rapid bootstrapping procedure (1000 pseudoreplicates) and the 'GTRGAMMA' model, followed by a maximum likelihood search using the best-fit models (as described above) for each partition.

2.5. Species trees

Our original intent was to utilize the faster coalescence of plastid nucleotide sequences, in conjunction with nuclear sequences, in order to improve species-level resolution. Our plastid gene tree, built over eight loci, provides the chloroplast lineage. For nuclear inference, we performed a BEST analysis using the three nuclear loci from 91 accessions. For comparison, the three nuclear loci were concatenated and used to infer a species tree from these same 91 accessions.

We identified six potential instances of chloroplast capture based on conflict between our plastid and nuclear gene trees and/or based on non-monophyly among plastid alleles in conjunction with outside information supporting putative hybridization (see Section 3.3). Post-speciation gene flow would violate an important underlying assumption of each of our species-tree models. The BEST model explicitly assumes that genes diverged before speciation, and thus that lineage sorting, not reticulate evolution, is the only source of discrepancies between gene trees and species trees (Liu et al., 2008). Likewise, hierarchical relationships are implicit in the Bayesian, parsimony, and likelihood models used for the gene trees and for the concatenated total-evidence analyses. As discussed above (see Section 1.2), introgressed plastid genes are more likely to become fixed in a population than nuclear loci. Further, plastid nucleotides are a significant component of this study: eight plastid loci (6910 bps) vs. three nuclear loci (2424 bps). To address the potentially distorting effects from plastid introgression, we identified several species which exhibit a plastid history that is significantly discordant from the nuclear gene trees (see Section 3.3) and then created a species 'compromise' by recoding their plastid nucleotides and plastid indel characters as missing data. For comparison, we used the same suppression of plastid nucleotides in the total-evidence analyses and for a nuclear-plastid BEST species tree.

All of the BEST analyses used MB-BEST vers. 2.2 (Liu, 2008; Liu and Pearl, 2007; Liu et al., 2008), an extension of MrBayes. This program applies a coalescent framework to estimate the posterior distribution of a species tree jointly with those of the gene trees for each locus, and has been successfully applied to several phylogenetic studies (Belfiore et al., 2008; Bossu and Near, 2009; Brumfield

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

et al., 2008; Linnen and Farrell, 2008; Spinks and Shaffer, 2009), although lack of convergence has been reported (Wares et al., 2009). This version of MB-BEST uses a uniform prior for the species tree and accounts for multiple accessions per species. Three partitions (and hence, three gene trees) were defined: (*ncpGS*; *NIA*; *PepC*). For the nuclear-plastid BEST, a fourth partition was defined for all non-suppressed plastid data. For each BEST data set, we used two simultaneous runs with 16 chains of MCMC for 20 million generations with a temperature parameter for chain mixing of 0.35. Based on the ratios of π between our loci (see Section 3.1), we elected to retain the default theta and gene mutation rate priors: thetapr = invgamma (3 and 0.003); genemupr = uniform (0.5 and 1.5). The optimum nucleotide substitution model was adopted for each locus, and the plastid partition was further defined as haploid. Numerous attempts failed to converge despite running up to 50 million generations (not shown). We were only able to achieve convergence toward a single optimum tree in either nuclear or nuclear-plastid data sets by setting a uniform prior for branch lengths (prset brlenspr = clock:uniform). Convergence was confirmed by at least two independent analyses. Each gene tree partition was evaluated as described above for Bayesian gene trees. Species cladograms were created by using a dummy matrix of the same length as the original data set, with one sequence per species, to summarize the species tree files created by BEST.

To test a total-evidence approach, we created a concatenated matrix with one record per species (32 ingroup tips) that combined nucleotide sequences for nrDNA (see below), and morphological characters (see below) with the nucleotide sequences and indel-coding characters for our eight plastid and three nuclear loci, yielding 10,694 total characters (Table 2). Our multiple-accession data set was collapsed by creating one consensus sequence for each species with variable nucleotide positions re-coded using IUPAC ambiguity codes. Thus, these species consensus records condensed the heterozygosity observed in low-copy nuclear loci as well as the variation observed across multiple individuals within each species. Variable indel coding positions were scored as missing data. Then, plastid data for the six species we identified with cytonuclear conflict (see Section 3.3) was re-coded as missing data as a compromise. For parsimony analyses, ambiguity coding was set to variable (i.e. pset mstaxa = polymorphic in PAUP*). For Bayesian analyses where a model to account for ambiguity coding is not available, we converted variable nucleotide positions to missing data. We concatenated one nrDNA sequence per species from a previous published study (Wagner et al., 2005): nucleotide sequences for *ITS* (GenBank accessions AY517655–AY517686) and *ETS* (GenBank accessions AY517688–AY517717). The *ETS* region for *S. laui* was sequenced for this study (GenBank accession GQ279733), but is missing for *S. nuttallii*. Outgroup sequences were available for *H. peploides* (*ITS*: AY517653 and AY517654; *ETS*: AY517687), *Minuartia* spp. (*ITS*: AY517679 and AY517650), and *Wilhelmsia physodes* (*ITS*: AY517652). Due to ambiguous alignment, we followed the submitting authors in excluding *ITS* positions 1–74 and 356–517 and *ETS* positions 1–25. We used 61 morphological characters described in Appendix 3 of the monograph (Wagner et al., 2005). Seven partitions were defined for the 32-allele matrix as described above for Bayesian analyses (Table 2). We used FigTree vers. 1.2.3 (Rambaut, 2010a) to modify all of our tree figures. Trees were outgroup rooted and the 'cartoon' option was used to collapse conspecific alleles. Each graphic file was exported from FigTree as a Windows Enhanced Metafile. Annotations and branch support graphics were added manually.

2.6. Dating

The plastid (rather than the nuclear) partition was selected to estimate divergence dates for several reasons. First, our plastid tree provides more resolved nodes than our nuclear gene trees (see Section 3.2). Second, modeling is simpler with a single partition that supports a single topology. Third, longer sequence lengths are an important factor in estimating ages (Brown and Yang, 2010) and our plastid data set contained many more nucleotides than our nuclear partition (6793 for eight plastid loci cf. 2424 for three nuclear loci; Table 2). Because reticulate evolution has affected *Schiedea*, this molecular clock represents how *plastid lineages*, not necessarily species, have diverged over time. Nonetheless, this is a useful exercise for *Schiedea* as the plastid lineages can be used to infer the age of the original colonization and, perhaps, the age of the crown radiation. As described above, individuals with putatively introgressed plastid genomes were excluded, but *S. spergulina* was retained in order to estimate the timing of plastid transfer. Ages were estimated using 'Bayesian evolutionary analysis by sampling trees' (BEAST) vers. 1.5.4 (Drummond and Rambaut, 2007; Heled and Drummond, 2009) with a simple HKY nucleotide substitution model, a UPGMA starting tree, a Yule tree prior (assuming constant speciation rate per lineage), 30 million generations, saving one tree every 1000 generations. Bayes factor comparisons in preliminary runs supported our selection of the uncorrelated relaxed lognormal rather than a strict clock model for this data set. Priors for the time to most recent common ancestor (TMRCA) were provided for two nodes to represent ages of the islands of Kauaì (4.7 Ma) and Oáhu (3.7 Ma; Clague, 1996). We tested three dates for the age prior on the root node. First, we set the root at 23 Ma to test the possibility that species could have island-hopped to newly formed islands dating back to the creation of Lisianski (Clague, 1996). We abandoned this calibration (and did not attempt the even older possibility of 30 Ma for the age of Kure) because we could not obtain resolved estimates for substitution rates among branches. Second, we assigned a prior for the root at 7.3 Ma to represent the age of Nihoa (Clague, 1996). This island, while old and eroded, currently supports one *Schiedea* species (*S. verticillata*). Third, we set a prior for the root at 4.7 Ma, the age of Kauaì. Runs were repeated with increasing numbers of generations and with parameter adjustments (e.g. window size) suggested by BEAST until the 'effective sample size' statistic was greater than 200 for each parameter. We used Tracer vers. 1.5 (distributed with the BEAST package) to summarize BEAST output, discarding 3 million generations as burn-in. One maximum clade credibility tree was created for each calibration scenario using TreeAnnotator vers. 1.5.2 (distributed with the BEAST package) with a 0.5 posterior probability limit, discarding 3000 trees as burn-in.

3. Results

3.1. Data matrices

The eight plastid loci had an aligned length of 6793 nucleotides and 117 indel codes for a total of 6910 characters, 329 of which are parsimony-informative (PI); the one-allele matrix for *ncpGS* had 845 nucleotides and 23 indel codes for a total of 868 characters (96 PI); *NIA* had 964 nucleotides and 101 indel codes for a total of 1065 characters (262 PI); and *PepC* had 615 nucleotides and 25 indel codes for a total of 640 characters (75 PI; Table 2). The two-allele data sets that were used to construct gene trees for nuclear loci yielded more PI characters than the one-allele matrices (Table 2). We did not detect genetic recombination in any of the tests for the three nuclear loci. We deposited 1284 new sequences in GenBank as accession numbers GQ223800–GQ225083. The nuclear BEST data set had 2424 nucleotides (363 PI) for 91-alleles, the total-evidence data set had 10,506 nucleotides, 128 indel-coding characters, and 61 morphological characters for a total of

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

10,695 characters (888 PI), while the nuclear-plastid BEST data set had 9217 nucleotides (624 PI) for 91 accessions (Table 2).

Nucleotide diversity for individual loci ranged from $\pi = 0.00425$ (SD = 0.00052) in *rpS16* to $\pi = 0.03278$ (SD = 0.00217) in *NIA* (Supplementary Table S2). The total nucleotide diversity across the eight plastid loci that we combined for a plastid gene tree was $\pi = 0.00640$ (SD = 0.00038). Based on the absence (or at least extreme rarity) of genetic recombination in circular plastid genomes, the lack of dramatic rate variation among our eight loci (Supplementary Table S2), the limited phylogenetic information in each plastid locus for these closely related species, and poor resolution in preliminary comparisons of phylogenies between individual plastid loci (data not shown), we chose to present analyses with these eight plastid loci concatenated as a single partition.

3.2. Gene trees

Parsimony searches for each of the four individual gene trees hit the maximum tree limit, returning 100,000 equally parsimonious trees. For Bayesian analyses, stationarity was generally reached early, and we discarded 2000 of 20,000 trees from each run as 'burn-in' except for the *PepC* gene tree, where we discarded 4000 trees.

For the plastid lineage, Bayesian and parsimony methods resolve essentially identical topologies (Fig. 1). Except for *S. globosa* and *S. kauaiensis* (see Section 3.3), the 91 accessions sampled in this study are largely (but not exclusively) resolved on the plastid gene tree into 32 clades that are consistent with species that have been recognized using a combination of morphological and nuclear characters (Wagner et al., 2005). In contrast, some of the slightly deeper nodes are poorly resolved on the plastid gene tree. Six of eight previously described taxonomic groups (see Plate 2 in Wagner et al., 2005) are recovered in the plastid gene tree: sections *Alphaschiedea*, *Alsinidendron*, and *Nothoschiedea* (Alp, Als, N; Fig. 1) as well as three monotypic sections: *Leucocalyx* (*S. attenuata*), *Polyneura* (*S. verticillata*) and *Anestioschiedea* (*S. apokremnos*). Sections *Schiedea* and *Mononeura*, which comprise the bulk of the species in the radiation, are polyphyletic on the plastid gene tree (S, MI, MII; Fig. 1).

For each nuclear gene tree, the topology of the parsimony strict consensus was similar to Bayesian estimates from the same data set (not shown). We chose to present the two-allele results from Bayesian analyses in order to show the full range of intraspecific variability (Supplementary Figs. S1–S3). Unlike the plastid gene tree, most species do not achieve complete monophyly on these trees. Paraphyly is particularly evident on the tips of the shallowest branches. However, the pattern on the nuclear gene trees, particularly within species assigned to sections *Schiedea* and *Mononeura*, is largely one of allele sharing among close relatives. Sections *Alphaschiedea, Alsinidendron*, and *Nothoschiedea* (and monotypic sections *Polyneura* (*S. verticillata*), and *Anestioschiedea* (*S. apokremnos*)) are resolved on each nuclear gene tree; monotypic section *Leucocalyx* (*S. attenuata*) is resolved on *NIA* and *PepC*, but not *ncpGS* trees. Section *Schiedea* is recovered on the *ncpGS* and *PepC* trees but is paraphyletic on the *NIA* gene tree. Conversely, section *Mononeura* is monophyletic in the *NIA* gene tree, but is polyphyletic on the other two gene trees. An early-branching clade, consisting of sections *Alsinidendron* and *Nothoschiedea*, *S. attenuata*, and *S. verticillata*, was recovered on the plastid, *ncpGS*, and *NIA* gene trees ('B' in Fig. 1, Supplementary Figs. S1 and S2). On the *PepC* gene tree (Supplementary Fig. S3), the B clade only differs in the placement of *S. verticillata* as a sister to *S. apokremnos* and in the placement of these species as sister to section *Schiedea* rather than an early-branching lineage. Previous evidence from nrDNA and morphology recovered a similar B clade, except for a reversed branching order between *S. attenuata* and *S. verticillata* (Wagner et al., 2005).

3.3. Cytonuclear discordance and identification of introgression events

There are four major conflicts between the plastid (Fig. 1) and nuclear gene trees (Supplementary Figs. S1–S3), plus two potential conflicts with weak or equivocal support. The first strongly supported case for chloroplast capture is the unequivocal placement on the plastid tree of three *S. spergulina* accessions (section *Schiedea*) within a clade of Kauaìan endemics belonging to section *Mononeura* (MII, Fig. 1). There is strong support for a sister relationship between *S. spergulina* and *S. stellarioides* plastid lineages. This relationship is dramatically different than the placement of *S. spergulina* within section *Schiedea* based on nuclear gene trees (Supplementary Figs. S1–S3) and in a previous phylogeny (Wagner et al., 2005). Based on the monophyly of *S. spergulina* on the plastid gene tree, supported by a moderate branch length, we can conclude that historical, rather than recent, introgression is highly likely (Sang and Zhong, 2000). However, the direction of the transfer (i.e. from section *Schiedea* to section *Mononeura* or vice versa) is equivocal.

The second example of strongly supported introgression is found in the *S. globosa* populations on Oáhu. *Schiedea globosa* has extant populations on Oáhu, Maui, Moloka'i, and Hawai'i, was formerly distributed on Lana'i, and has the widest distribution of any species in the lineage (Wagner et al., 2005). On our plastid gene tree, three *S. globosa* individuals from Oáhu resolve strongly supported as sister to a rare Oáhu endemic, *S. adamantis*, while the other two accessions are strongly supported as closer to all other members of section *Schiedea* (Fig. 1). On the *ncpGS* gene tree (Supplementary Fig. S1), one of the *S. globosa* Oáhu) accessions is sister to one of the *S. adamantis* accessions, although without branch support. In the other nuclear gene trees (Supplementary Figs. S2 and S3), all five *S. globosa* accessions are monophyletic, and are not part of the clade containing *S. adamantis*. Based on the topology of the plastid gene tree (Fig. 1), the plastid lineage found in *S. globosa* accessions from Oáhu appears to be derived from *S. adamantis*.

The third strongly supported case of plastid introgression involves two closely related but morphologically distinct members of section *Schiedea*: *S. lydgatei* and *S. sarmentosa.* Our plastid gene tree resolves a clade containing all six accessions of both species (Fig. 1). One *S. sarmentosa* allele joins *S. lydgatei* on the *ncpGS* gene tree, but not on the *NIA* or *PepC* gene trees (Supplementary Figs. S1–S3). With the evidence at hand, the introgression of nuclear genes and the direction of the transfer remain ambiguous. We arbitrarily selected *S. lydgatei* to suppress plastid information in the compromise data set.

The fourth strongly supported case of plastid introgression is the monophyletic lineage shared by all five of our *S. hookeri* and *S. mannii* accessions (Fig. 1). Nuclear alleles from these same accessions do not cluster (Supplementary Figs. S1–S3), congruent with findings from a previous phylogeny (Wagner et al., 2005). The source populations for two of our *S. hookeri* collections are sympatric with *S. mannii* (between Wai'anae and Makaha valleys on Oáhu), although the third *S. hookeri* collection (Kaluakauila gulch) is ca. 2 km from the nearest documented *S. mannii* populations on 'Ohikilolo Ridge (Wagner et al., 1995). Our results suggest one or more historic hybridization events that lead to a single plastid lineage persisting in both species. The lack of resolution on the plastid gene tree does not clarify which parental chloroplast persisted, so our decision to suppress the plastid sequences for *S. mannii* was arbitrary.

More obscure evidence for chloroplast introgression was observed for *S. kauaiensis* (section *Mononeura*), which is paraphyletic

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

Fig. 1. Bayesian majority-rule gene tree using nucleotide sequences and indel-coding characters for eight plastid loci in 92 accessions (Table 2). Thick branches are PP $\geqslant$ 90 and parsimony BS $\geqslant$ 70. Conspecific alleles collapsed into triangular cartoons. For polyphyletic species, the number of individuals is in parenthesis. Asterisks mark plastid data that was suppressed in the compromise data sets (see Section 3.3). Clade abbreviations: Alp, *Alphaschiedea*; Als, *Alsinidendron*; B, an early-branching clade; MI, *Mononeura* I; MII, *Mononeura* II; N, *Nothoschiedea*; S, section *Schiedea*.

on the plastid gene tree (Fig. 1). One accession from the Mahanaloa Valley, Kauaì, (KW7430) nests within an unresolved clade containing *S. helleri* and *S. membranacea* (section *Alphaschiedea*). On the *ncpGS* gene tree (Supplementary Fig. S1), *S. kauaiensis* is also paraphyletic, but in this case the KW7430 accession is the only one that resolves in the expected position as sister to the morphologically similar *S. stellarioides*. *Schiedea kauaiensis* is also paraphyletic on the *NIA* gene tree, but in yet a different pattern (with four of the six alleles monophyletic and a single allele clustering with its close relative and morphologically similar *S. perlmanii*, and a single allele clustering with alleles from *S. jacobii* and *S. kaalae* in the MI clade (Supplementary Figs. S2). The *PepC* alleles for *S. kauaiensis* form an unresolved polytomy with the monophyletic *S. perlmanii* alleles in the *PepC* gene tree (Supplementary Fig. S3). Thus, *S. kauaiensis* exhibits characteristics that could be explained by either retention of ancestral polymorphism or introgression. Because of the possibility that the paraphyletic plastid accession (KW7430) may be due to chloroplast capture, our compromise excluded its plastid data. However, the direction of the transfer is equivocal. The placement of *Alphaschiedea* on the plastid tree (Fig. 1; sister to MII species that are sympatric on the island of Kaua'i) rather than as an early-branching lineage (Wagner et al., 2005) and on the nuclear

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

BEST tree (see Section 3.4) suggests that *S. helleri* and *S. membranacea*, the only two extant species of *Alphaschiedea*, may have acquired a chloroplast lineage from the younger *Mononeura* group.

Although a previous study (Soltis et al., 1996) reported potential chloroplast transfer between *S. ligustrina* and *S. kealiae*, we observed only weak evidence for introgression. On our plastid gene tree, (Fig. 1), there is very weak support for their sibling status (BS < 50; PP = 55; depicted as unresolved in Fig. 1). The three accessions of *S. kealiae* are not polyphyletic on the *PepC* gene tree (Supplementary Fig. S3), but no sister relationship with *S. ligustrina* is supported. *Schiedea kealiae* alleles are unresolved on the *ncpGS* and *NIA* gene trees (Supplementary Figs. S1 and S2). Because our results are not inconsistent with evidence for plastid introgression (Soltis et al., 1996), we chose to exclude the plastid information for *S. kealiae* in our compromise data set.

Based on these cytonuclear conflicts, we re-coded the plastid partition as missing data in the compromise data set for the following accessions: (i) all accessions of *S. spergulina*; (ii) all accessions of *S. mannii*; (iii) all accessions of *S. lydgatei*; (iv) three *S. globosa* accessions from Oáhu; (v) one accession of *S. kauaiensis*, and (vi) and all accessions of *S. kealiae* (asterisks, Fig. 1). This compromise scenario was used for the total-evidence data set and to estimate a nuclear-plastid species tree using BEST.

3.4. Species trees

Although the nuclear BEST tree (Fig. 2), the nuclear concatenated tree (Fig. 3), the total-evidence concatenated trees (Fig. 4a, 4b, 4c), and the nuclear-plastid BEST tree (Fig. 4d) each resolve similar major clades (Alp, Als, B, MI, MII, N, and S), they differ in their placement. The parsimony analysis is unique among the total-evidence trees in its placement of *Alphaschiedea* in a basal position (Fig. 4a) in agreement with the nuclear BEST and nuclear concatenated trees (Figs. 2 and 3). The nuclear BEST and nuclear concatenated trees are also unique in resolving a monophyletic section *Mononeura*. The nuclear BEST tree (Fig. 2) strongly supports *S. spergulina* as sister to the remaining section *Schiedea*. The nuclear concatenated tree (Fig. 3) is less resolved

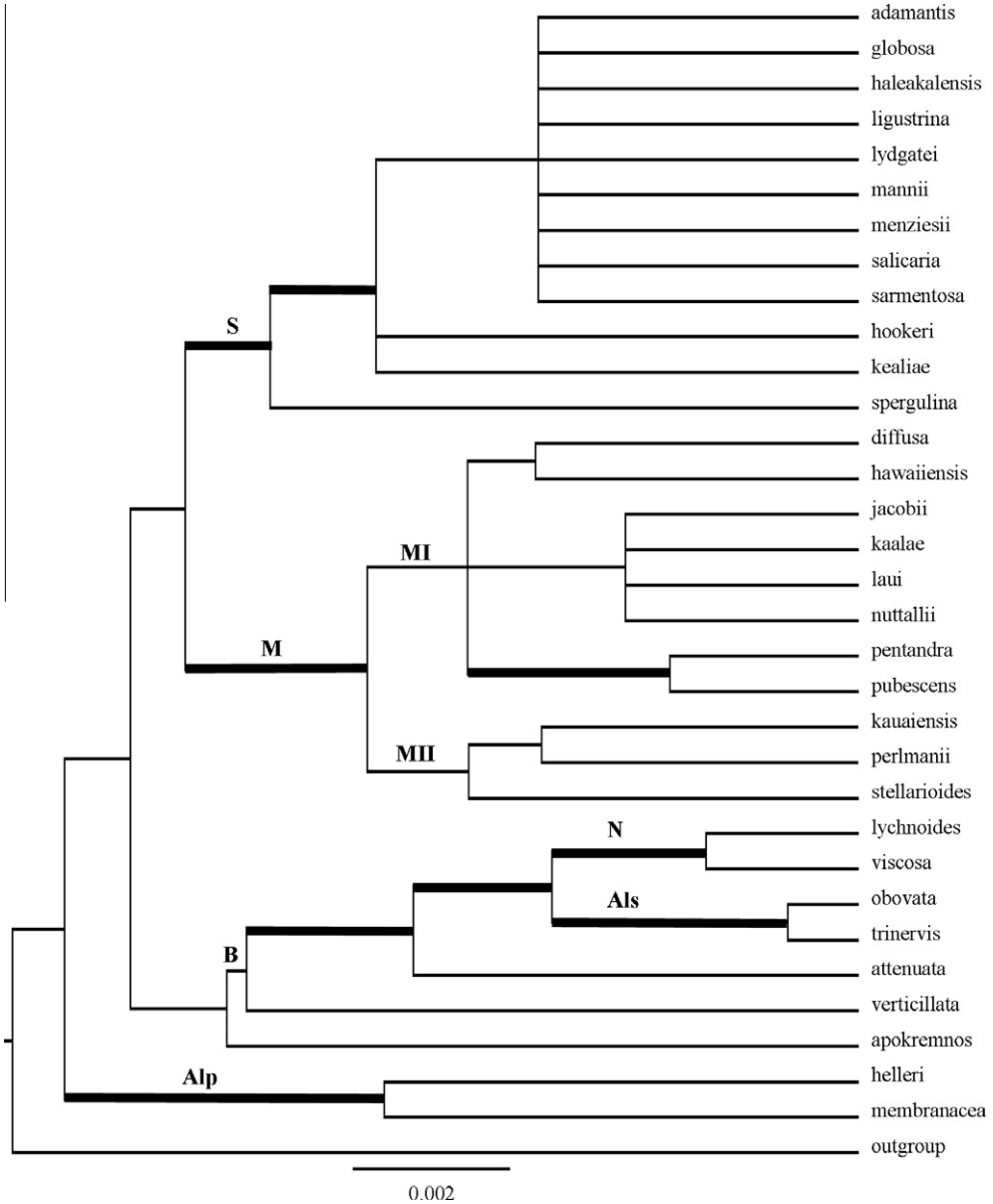

Fig. 2. Nuclear BEST species tree using three nuclear loci (Table 2). Thick branches are PP ⩾ 90. Clade abbreviations as described in Fig. 1, plus M, *Mononeura*.

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

Fig. 3. Bayesian majority-rule nuclear concatenated species tree using nucleotide sequences in 92 accessions (Table 2). Thick branches are PP ⩾ 90 and parsimony BS ⩾ 70. Conspecific alleles collapsed into triangular cartoons. For polyphyletic species, the number of individuals is in parenthesis. Clade abbreviations: Alp, *Alphaschiedea*; Als, *Alsinidendron*; B, an early-branching clade; M, *Mononeura*; N, *Nothoschiedea*; S, section *Schiedea*.

than the nuclear BEST tree (Fig. 2) at deeper nodes corresponding to taxonomic sections. For example, section *Schiedea* resolves as a grade (Fig. 3). In contrast, the nuclear concatenated species tree resolves several paraphyletic nodes that are not supported by other methods (thick lines indicate PP ⩾ 90 in Fig. 3).

3.5. Genealogical sorting Indices

For the plastid gene tree, nineteen species are monophyletic (*gsi* = 1.0; Supplementary Fig. S4a). The nuclear gene trees have fewer monophyletic species: *ncpGS* has eleven (Supplementary Fig. S4b), *NIA* has fifteen (Supplementary Fig. S4c), and *PepC* has eight (Supplementary Figs. S4d). Although twelve species have *gsi* < 0.5 on the plastid gene tree, the *gsi* for all 32 *Schiedea* species is statistically significant. In contrast, each nuclear gene tree reveals several species that lack significant *gsi* results (Supplementary Figs. S4b–d, unfilled square symbols). Three species (*S. diffusa*, *S. verticillata*, and *S. viscosa*) have a *gsi* = 1.0 in every gene tree. At the other end of the scale, six species (*S. hookeri*, *S. kaalae*, *S. kauaiensis*, *S. ligustrina*, *S. lydgatei*, and *S. mannii*) have *gsi* < 0.5 in every gene tree. Although fourteen species have *gsi* < 0.5, the ensemble statistic estimates a significant *gsi* for every species (Fig. 5).

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

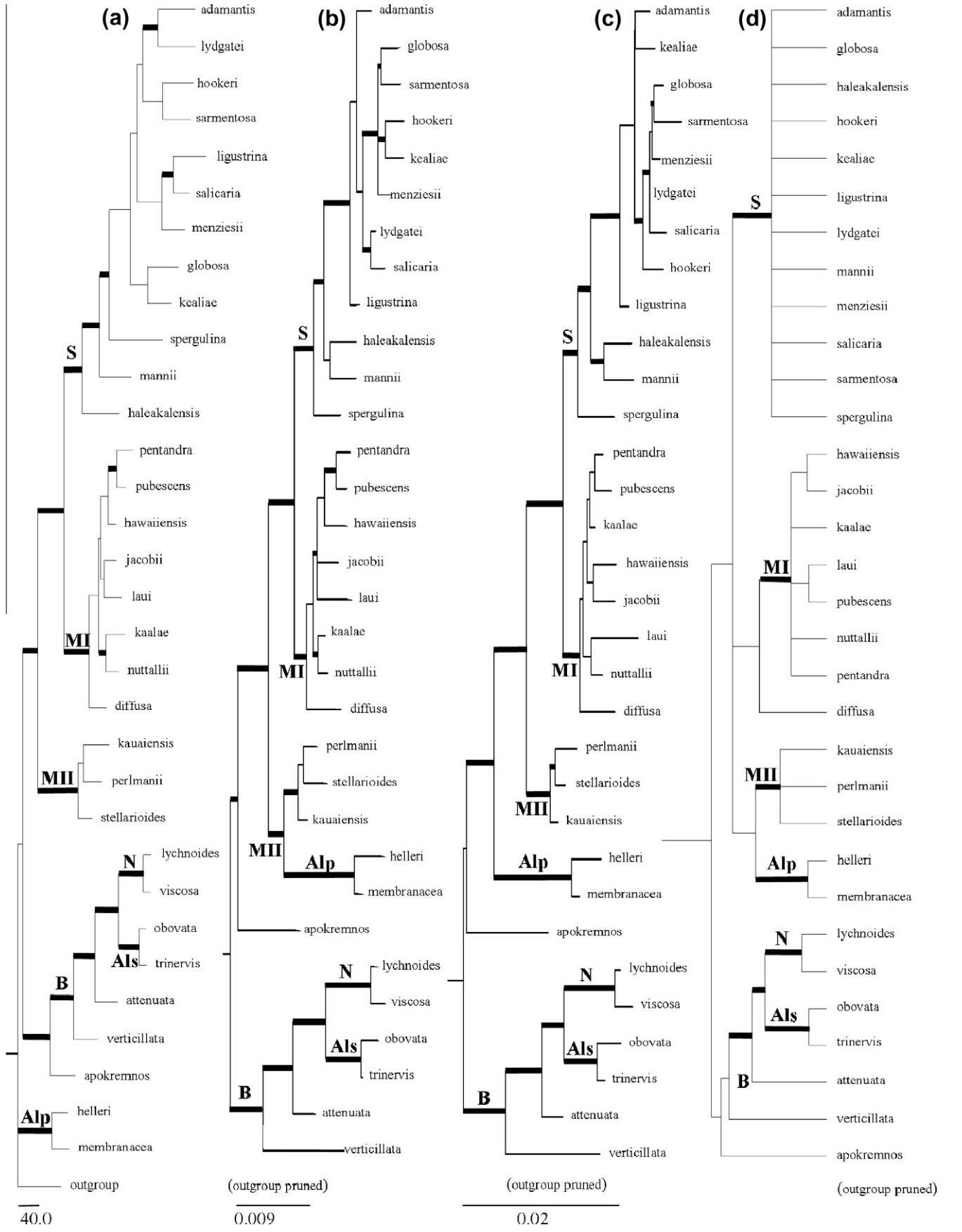

Fig. 4. *Schiedea* species trees based on compromise data sets that suppress plastid sequences with putative introgression (see Fig. 1; Table 2): (a) 32-allele total-evidence: parsimony (single shortest tree); (b) 32-allele total-evidence: Bayesian majority-rule; (c) 32-allele total-evidence: maximum likelihood; (d) 91-allele nuclear-plastid BEST species tree. Thick branches are BS $\geqslant$ 70 or PP $\geqslant$ 90. Clade abbreviations as described in Fig. 1.

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

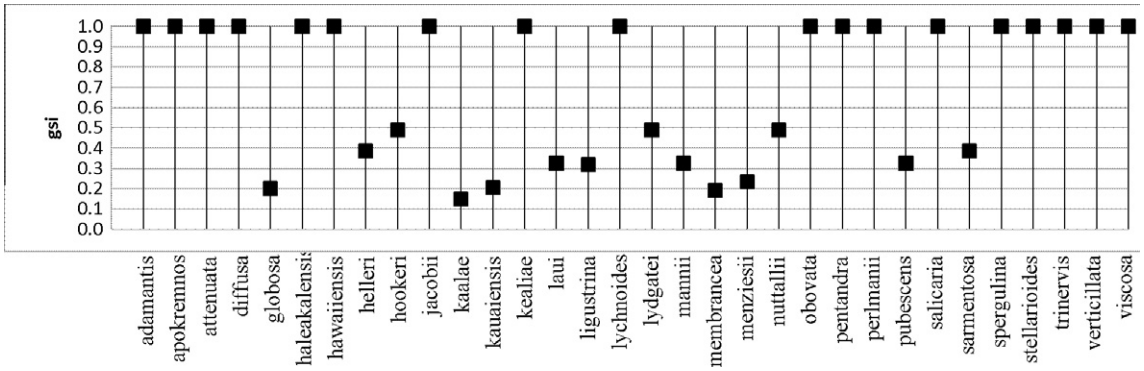

Fig. 5. Comparison of the degree of exclusive ancestry (*gsi*; genealogical sorting index) among accessions of each *Schiedea* species. The ensemble statistic weights the *gsi* from each of the four gene trees with its probability. A *gsi* of 1.0 indicates monophyly. Filled squares denote significant indices at $P = 0.05$ level. See Supplementary Fig. S4 for the *gsi* of each independent gene.

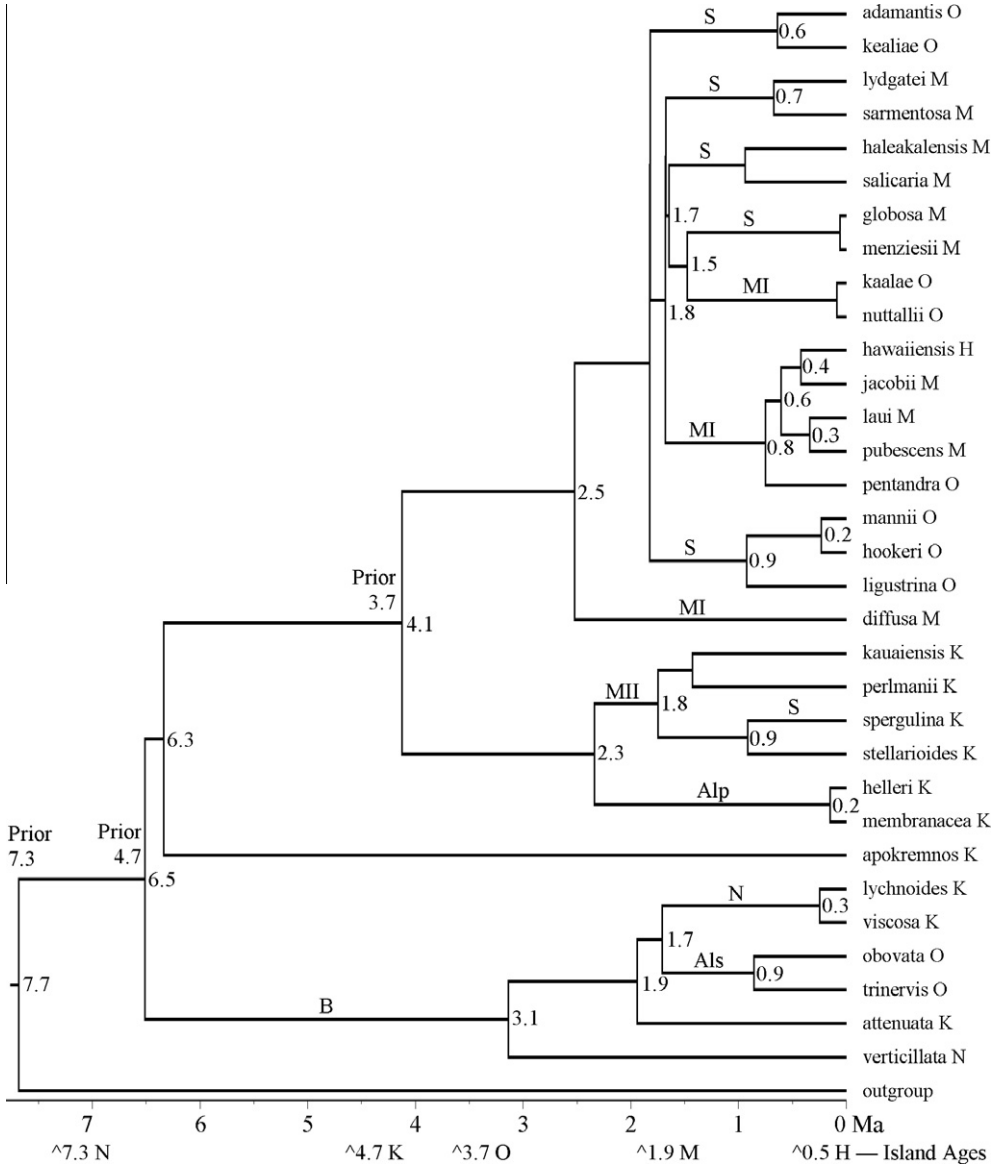

Fig. 6. Plastid phylogeny drawn proportional to median ages inferred from BEAST. Data set was reduced (see Section 3.3 and Fig. 1) to exclude putative introgression, except that *Schiedea spergulina* was included here. Three prior parameters were specified in BEAST for the 'time to most recent common ancestor' (TMRCA): the age of Nihoa (7.3 Ma) for the root, the age of the island of Kauaì (4.7 Ma), and the age of the island of Oáhu (3.7 Ma) as indicated on the tree. For nodes with $\geqslant 0.5$ posterior probability, the median inferred age is shown. Supplementary Table S3 presents the 95% highest posterior density intervals for these age estimates. Geographic source of accessions used is abbreviated next to taxon name: N, Nihoa; K, Kauaì; O, Oáhu; M, Maui Nui; H, Hawai'i. Clade abbreviations as described in Fig. 1.

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

3.6. Node ages and substitution rates

With priors of 7.3 Ma and 4.7 Ma for TMRCA of the root, we obtained mean plastid rates of 1.02×10^{-9} substitutions per site per year (SE = 4.54×10^{-12}) and 1.40×10^{-9} substitutions per site per year (SE = 6.42×10^{-12}), respectively. In the 7.3 Ma calibration scenario, the root and the deepest two nodes have median ages that predate the age of Kauaì, the oldest current high island (Fig. 6). This would suggest that *Schiedea* plastid lineages might have arisen on older, currently eroded islands. However, 95% confidence intervals (Supplementary Table S3) are wide, allowing the possibility that the crown radiation arose after Kauaì was above water. In the 4.7 Ma root calibration scenario, median ages estimated for all nodes representing inter-island divergences are more recent than the age of the currently inhabited island (Fig. 7; Supplementary Table S3). The TMRCA for section *Schiedea* is estimated at 1.3 Ma using the 4.7 Ma calibration scenario (Fig. 7). For the 7.7 Ma calibration scenario, the age of the TMRCA for section *Schiedea* is unclear, falling between nodes aged at 2.5 Ma and 1.8 Ma (Fig. 6). Both estimations suggest a surprisingly ancient origin for the 12 closely related species in section *Schiedea*.

4. Discussion

4.1. Species monophyly

By the criterion of a significant ensemble genealogical sorting index (*gsi*), all 32 extant taxonomic species of *Schiedea* are monophyletic and therefore 'good' species based on the gene trees presented here. The *gsi* calculated from four individual gene trees, as well as the ensemble *gsi* calculated by weighting indices by their probability on each tree, suggest a high degree of common ancestry for named taxonomic species of *Schiedea*. The weighting scheme provided in the ensemble *gsi* statistic appears to be a useful way to incorporate the uncertainty in heterogeneous gene trees. For example, *S. adamantis*, *S. kealiae*, *S. pentandra*, *S. salicaria*, and *S.*

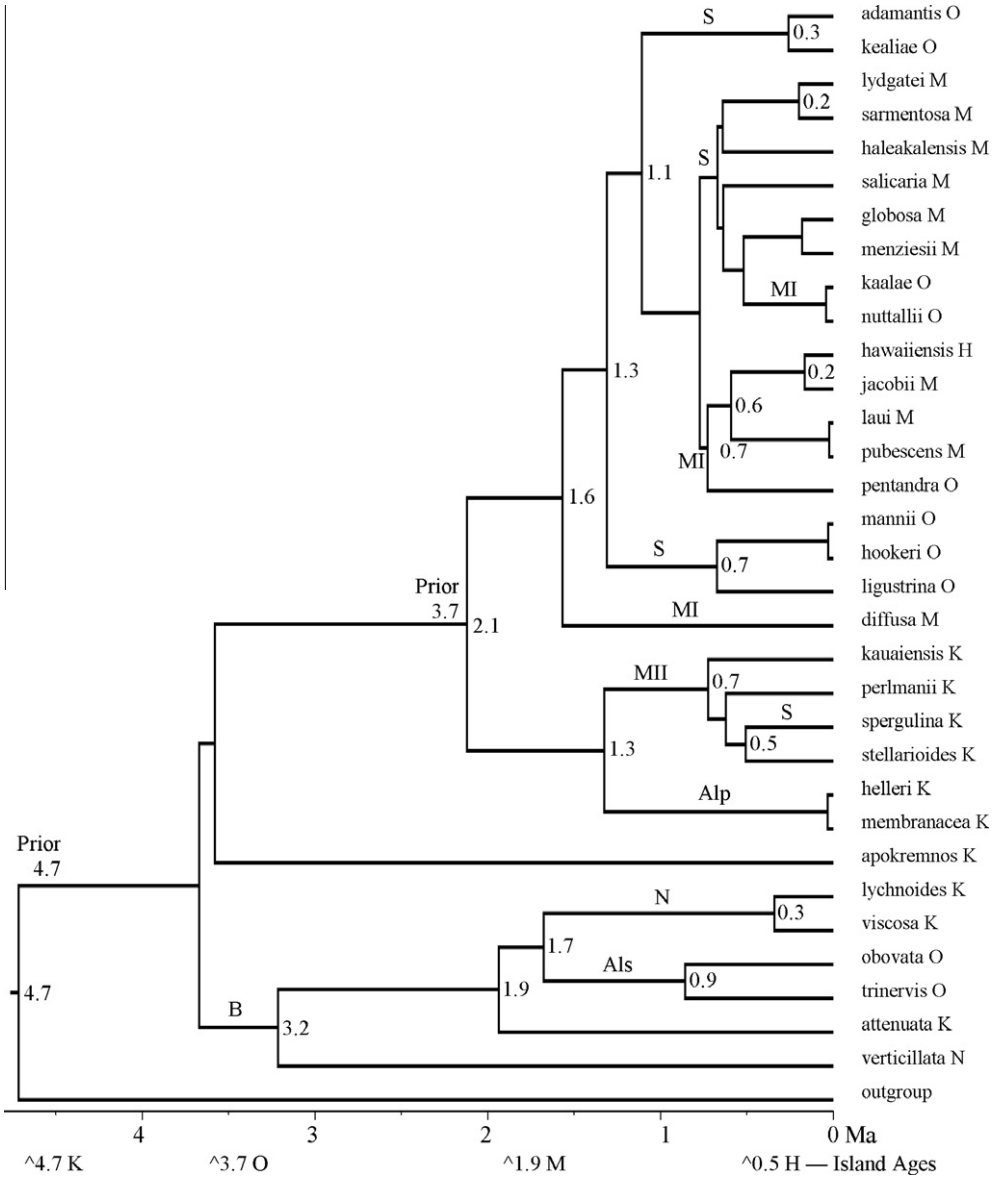

Fig. 7. Plastid phylogeny drawn proportional to median ages inferred from BEAST. Data set was reduced (see Section 3.3 and Fig. 1) to exclude putative introgression, except that *Schiedea spergulina* was included here. Two prior parameters were specified in BEAST for the 'time to most recent common ancestor' (TMRCA): the age of the island of Kauaì (4.7 Ma) for the root and the age of the island of Oáhu (3.7 Ma) as indicated on the tree. For nodes with $\geqslant 0.5$ posterior probability, the median inferred age is shown. Supplementary Table S3 presents the 95% highest posterior density intervals for these age estimates. Geographic source of accessions as described in Fig. 6; clade abbreviations as described in Fig. 1.

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

spergulina each receive a non-significant *gsi* (Supplementary Fig. S4) on one gene tree that can be attributed to a lack of resolution rather than support for polyphyly. The high ensemble statistic for each of these species (*gsi* = 1.0) results from their higher probabilities for monophyly on the other gene trees. This appears to be a better reflection of species monophyly than might be apparent from an independent evaluation of each poorly resolved gene tree. Thus, a comparison of *gsi* scores among partitions may offer an efficient method to screen for potential reticulate evolution in large data sets.

Despite strong cytonuclear conflict, it appears that plastid introgression in *Schiedea* is mainly historical, and that putatively introgressed plastid lineages have become fixed in most (but not all) species that exhibit evidence of past introgression. In fact, only two species (*S. globosa* and *S. kauaiensis*; discussed below) display plastid paraphyly. Coalescence times for plastid alleles are expected to be four times faster than for nuclear alleles, and given likely relatively small ancestral effective population size (for rare, insular species with small population sizes), strong species monophyly for plastid alleles in Hawaiian *Schiedea* is not surprising.

Strong species monophyly in *Schiedea* is likely fostered by extreme topographic variation in the Hawaiian Islands, leading to allopatric divergence of populations and species. Wallace et al. (2009) reported convincing evidence for allopatric divergence of populations, potentially leading to incipient speciation among populations of the most widespread species in the lineage, *S. globosa*. Additionally, many closely related *Schiedea* species are allopatric and occur on different islands. Several morphologically similar species groups are characterized by strong allopatry. For example, within the closely related species of section *Schiedea*, *S. ligustrina* and *S. adamantis* bear close morphological resemblance, but currently occur on different mountain ranges on Oáhu. *Schiedea lydgatei* and *S. salicaria* occur on different islands now and different parts of the Maui Nui complex historically. *Schiedea kealiae*, *S. menziesii*, *S. sarmentosa* and *S. hookeri* occur on different islands or on different parts of the Waiànae Mountain range on Oáhu, on different islands, or on different parts of Maui Nui historically.

In addition to allopatric divergence of *Schiedea* populations and species, interspecific gene flow may have been prevented in some cases by differences in breeding systems among several groups of sympatric or potentially sympatric species. Examples include *S. obovata* (a facultatively autogamous hermaphrodite) and *S. nuttallii* (an outcrossing hermaphrodite) which occur in sympatry in the Waiànae Mountains of Oáhu, and *S. viscosa* (facultatively autogamous hermaphrodite) and *S. stellarioides* (outcrossing hermaphrodite with a relatively high selfing rate) which occur in sympatry in the Waiàlae Falls area of the central highlands on Kauaì. When naturally occurring morphological intermediates have been found (e.g. between *S. lydgatei* and *S. sarmentosa* on Molokaʻi, or *S. salicaria* and *S. menziesii* on west Maui), the reproductive systems are at least partially outcrossing in both species (Weller et al., 2001). As a result, differences in breeding systems among different species of *Schiedea* may also have played a role in maintaining species boundaries, and contributed to strong monophyly of named taxonomic species.

4.2. Hybridization, introgression, and reticulate evolution in Schiedea

Evidence for species monophyly notwithstanding, our results strongly support scenarios of historical hybridization within the *Schiedea* radiation, observed as cytonuclear conflicts that occur at both deep and shallow nodes in the tree. Because of these well-supported chloroplast capture events, we did not combine nuclear and plastid data for species tree inference except when attempting the total-evidence and nuclear-plastid BEST analyses. Additional evidence for introgression is also present, but less well supported. Once multiple loci were sampled and gene trees compared, it appeared to be relatively straightforward to identify both major cases of introgression (i.e. those involving species not known to be sister species, and in some cases occupying very divergent positions among different sections/lineages) and those with only weak or equivocal support for introgression of the plastid genome. In all of these instances, the plastid partition was inferred to be introgressed or potentially introgressed. In three of six cases (*S. sarmentosa*, *S. kauaiensis* and *S. kealiae*), our results were equivocal in also supporting introgression of nuclear alleles. We chose to leave those equivocal nuclear alleles in all of our species tree estimations because: (a) the nuclear genome has a four times slower rate of coalescence than the plastid genome, supporting the notion that lack of strong support for monophyly of nuclear alleles in those three species is more likely due to retention of ancestral polymorphisms than introgression; and (b) recent studies have indicated that organelle genes are often preferentially introgressed across species boundaries due to reduced gene flow incurred as a result of their uniparental inheritance (Currat et al., 2008). Indeed, organellar introgression is supported exclusively in three of the six cases (*S. spergulina*, *S. globosa* on Oáhu and *S. mannii*). Currat et al. (2008) demonstrated that chloroplast transfer typically occurs from a local species to an invading species. Given the degree to which invasions in the form of colonization of new islands and of new areas on islands are likely to have figured in the history of this insular lineage, our observed pattern of predominantly organellar introgression seems reasonable.

4.2.1. S. spergulina and S. stellarioides

These two species are partially sympatric, although no morphological intermediates have been reported (Weller et al., 2001). *Schiedea stellarioides*, presumed extinct until rediscovered in 1991, is genetically depauperate (Weller et al., 1996) and only two extant populations are known. Both species likely had larger distributions in the past, with potential for extensive areas of sympatry. We note that all three of our *S. spergulina* accessions are from the Waimea Canyon; the populations from Lawai (*S. spergulina* var. *leiopoda* Sherff) were not sampled in this study. Because *S. spergulina* is the only member of section *Schiedea* that inhabits the oldest island of Kauaì, its relationship to other species in section *Schiedea* is critical to understanding the direction of plastid introgression (Fig. 1). If *S. spergulina* represents a back-colonization from ancestors that inhabited one of the younger islands, it may have obtained a plastid lineage from *S. stellarioides*, perhaps while conspecifics were still rare. This would fit a pattern of an invading species acquiring the organelle lineage of a local species (Currat et al., 2008). Alternatively, if *S. spergulina* is sister to all other species in section *Schiedea*, as strongly supported by our nuclear BEST tree (Fig. 2), by our Bayesian and ML total-evidence trees (Fig. 4b, 4c), and weakly supported by our nuclear concatenated tree (Fig. 3), then it may represent the sole progenitor of section *Schiedea* that has persisted on Kauaì. This implies that the hybridization event may have occurred concurrent with the onset of the rapid radiation of section *Schiedea*, a lineage containing all but one of the florally dimorphic *Schiedea* species. Interestingly, *S. spergulina* is dioecious. Its basal position within section *Schiedea* suggests dimorphism occurred early in section *Schiedea*. Alternatively, the occurrence of gynodioecy in *S. apokremnos*, one of the earliest diverging species in the lineage (Fig. 2), suggests that dimorphism may have been present in ancestors of the section *Schiedea* lineage.

4.2.2. S. globosa and S. adamantis

Schiedea globosa reveals the most disparate gene trees of any species in our study: complete monophyly in *ncpGS* and *PepC*, in contrast to low (but significant) *gsi* on the plastid and *NIA* gene

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

trees. In conjunction with gene tree incongruence patterns, this adds to the evidence that introgression has affected *S. globosa*. Populations of *S. globosa* follow the progression rule for island biogeography (Funk and Wagner, 1995), such that *S. globosa* populations on the older island of Oáhu appear to represent the ancestral genotypes (Wallace et al., 2009). Based on the prediction that organelle introgression tends to occur from a local species into an invading species (Currat et al., 2008), the asymmetric introgression of the plastid genome from *S. adamantis* into *S. globosa* would suggest that *S. globosa* invaded the range of the former species on Oáhu after the colonization of the younger islands.

4.2.3. S. lydgatei and S. sarmentosa

Putative hybrids between these two species are intriguing because *S. lydgatei*, a shrub with diffuse inflorescences that grows in dry shrublands, was suggested to have reverted to a hermaphroditic breeding system from a gynodioecious ancestor (Norman et al., 1995, 1997). *Schiedea sarmentosa* is a gynodioecious perennial with contracted inflorescences that grows on rocky cliffs. While numerous morphological intermediates have been reported where these species' habitats meet (Wagner et al., 2005), our results infer that introgression may be largely limited to the plastid lineage. Incomplete lineage sorting likely obscures the pattern of introgression in the nuclear genomes of these closely related siblings.

4.2.4. Alphaschiedea and Mononeura

The direction of introgression for one important cytonuclear conflict was not obvious in a comparison of the plastid (Fig. 1) and three nuclear gene trees (Supplementary Figs. S1–S3). On the total-evidence trees, *Alphaschiedea* is placed in a basal position by parsimony (Alp; Fig. 4a) but not Bayesian, ML or BEST methods (Fig. 4b, 4c, 4d). This conflict might be explained by long branch attraction in parsimony. What is more, the paraphyletic placement of *S. kauaiensis* (see Section 3.3) suggests that the KW7430 accession acquired a plastid lineage from *Alphaschiedea*. However, the nuclear BEST tree (Fig. 2) and the nuclear concatenated tree (Fig. 3) recover a basal position for *Alphaschiedea*, in agreement with previous nrDNA and morphological trees (Wagner et al., 2005). Cytonuclear conflict between the plastid tree and the nuclear BEST tree (Fig. 1 vs. Fig. 2) suggests that plastid introgression occurred in the opposite direction: i.e. a *Mononeura*-style plastid into the rather divergent *Alphaschiedea* lineage. This deep discordance was unexpected, and our decision to suppress the paraphyletic *S. kauaiensis* plastid allele rather than the *Alphaschiedea* alleles makes our total-evidence and nuclear-plastid BEST trees positively misleading.

4.2.5. Effects of Introgression on phylogeny

While most of the cases of plastid introgression detailed above appear relatively straightforward to characterize, introgression of nuclear alleles cannot be discounted in several instances. The majority of the nuclear conflicts are among very closely related species within a single taxonomic section. Several recent studies concluded that BEST may be fairly robust to potential violations of the assumption that no gene flow occurred after speciation. Belfiore et al. (2008) suggested that narrow hybrid zones may form between sympatric *Thomomys* subgenus *Megascapheus* pocket gopher species, but determined that population structure combined with low dispersibility and low density populations have contributed to rapid genetic drift and hence, "cohesiveness" of taxa within the lineage. Brumfield et al. (2008) and Liu et al. (2008) have reported low levels of introgression among closely related or sister species of manakins and suggested that hybridization may not affect the results of their estimation of species trees using BEST when it has occurred between very closely related species, affecting branch length rather than topology for those sister species.

Methods to incorporate models of reticulate evolution in algorithms for species tree estimation are recognized as important goals (Edwards et al., 2007), reflecting the notion that reticulate events, as well as cladistic events, contribute to the history of a group (Moore et al., 1995). The possibility remains that additional *Schiedea* plastid lineages that are not clearly identified as introgressed are tracking geography rather than hierarchical events. Further, the actual extent of introgression in the *Schiedea* nuclear genome remains unknown and the relative sensitivity of the BEST method to increasing levels of violation also remains untested.

4.3. Schiedea species trees

Simulations by Maddison and Knowles (2006) demonstrated that three nuclear loci may be sufficient for inferring a species tree for a recently diverged group when three individuals per species are sampled. However, our 91 accession data set far exceeds their largest simulation (27 individuals in each of eight species). Further, ours is among the largest data set analyzed to date using the BEST method. To our knowledge, only Linnen and Farrell (2008) analyzed more samples. Although we had to use a uniform clock prior in BEST in order to achieve convergence, the results appear to be stable based on multiple runs. The BEST *Schiedea* species tree (Fig. 2) provides a better resolution at the sectional level than the nuclear concatenated tree (Fig. 3), but despite utilizing 2424 characters from three nuclear loci, infers a nine-species rake within section *Schiedea* and a four-species rake within section *Mononeura* (Fig. 2). Finer relationships supported on the concatenated nuclear species tree for these sections may have merit, but they are difficult to interpret for a species hypothesis because this tree resolves accessions of *S. kealiae*, *S. ligustrina*, *S. lydgatei*, *S. mannii*, and *S. salicaria* as polyphyletic (Fig. 3). The relatively low *gsi* indices for these five species (Fig. 5), does not conflict with their polyphyletic placement on the nuclear concatenated tree.

We also tested the BEST species tree for robustness by comparing it with (1) a nuclear-plastid BEST species tree; and (2) a total-evidence method that incorporates information from nuclear, plastid, nrDNA, and morphological partitions into a single species-tree analysis. Due to our observation of chloroplast capture, these latter two methods used a compromise data set to suppress some plastid data (see Section 3.3). While it is not possible to directly compare the results of the total-evidence and nuclear-plastid BEST analyses because different data sets were used, the most striking comparison is that all of the concatenated total-evidence species trees are much more resolved than the nuclear-plastid BEST tree. These trees also display conflicting resolutions that can be attributed to analytical method (Fig. 4a–c). A few of the conflicts are strongly supported, even among terminal taxa (also see Section 4.5). On the other hand, the total-evidence and nuclear-plastid BEST trees agree with each other and with nuclear-only species trees on some aspects of the topology, particularly for deeper taxonomic sections. We found consistent support for monophyletic sections *Alphaschiedea*, *Alsinidendron*, and *Nothoschiedea* described by Wagner et al. (2005). A monophyletic section *Schiedea* is recovered on all species trees except the nuclear concatenated tree, where a section *Schiedea* grade can be attributed to lack of resolution rather than support for paraphyly (Fig. 3).

The divergent plastid lineage exerts an influence beyond the identified examples of cytonuclear conflict. First, all of the total-evidence trees, the nuclear-plastid BEST species tree, and the plastid tree (Fig. 1) resolve section *Mononeura* as paraphyletic with section *Schiedea* (MI, MII; Fig. 4a–d). In contrast, section *Mononeura* is monophyletic on both nuclear species trees (Figs. 2 and 3) and on a smaller data set of nrDNA and morphology (Wagner et al., 2005). This

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

suggests a historical chloroplast transfer of the plastid lineages between sections *Schiedea* and *Mononeura* that is creating this paraphyletic resolution in every data set where plastid nucleotides are present. Similarly, both nuclear species trees reveal a basal resolution of *Alphaschiedea* (Figs. 2 and 3), even though *NIA* (Supplementary Fig. S2) is the only gene tree with this topology. This placement for *Alphaschiedea* agrees with previous reports (Wagner et al., 2005). For the total-evidence species trees, only parsimony (Fig. 4a) recovers this placement. Notably, our plastid tree resolves *Alphaschiedea* sister to the MII clade (Fig. 1). The simplest explanation for these results is that one of our compromise assumptions (that the plastid lineage from *Alphaschiedea* had been transferred to *S. kauaiensis*) is incorrect. Despite ambiguous placement of *Alphaschiedea* on individual nuclear gene trees (Supplementary Figs. S1–S3), the nuclear species trees (Figs. 2 and 3) clearly resolve the lineage as expected from nrDNA and morphology. Therefore, the placement of *Alphaschiedea* in our compromise scenarios appears to be driven by its plastid lineage. Further, both nuclear species trees (Figs. 2 and 3) resolve *S. diffusa* and *S. hawaiiensis* as sister taxa. The resolution of *S. diffusa* as sister to remaining *Mononeura* on the total-evidence and nuclear-plastid BEST trees (Figs. 4a–d) also appears to be driven by the topology of the plastid lineage (Fig. 1).

Schiedea apokremnos (monotypic section *Anestioschiedea*) is resolved as sister to the B clade in the nuclear BEST, total-evidence parsimony, and nuclear-plastid BEST species trees (Figs. 2 and 4a and d). The plastid tree (Fig. 1) places *S. apokremnos* sister to members of sections *Schiedea* and *Mononeura,* resembling a previous nrDNA and morphology tree (Wagner et al., 2005). *Schiedea apokremnos* is known only from valleys and ridges of the Napali Coast on northwestern Kauaì, the oldest part of the oldest current high island in the chain. In this case, our improved nuclear data set clearly places this species among the basal lineages of the radiation, consistent with its geographical location.

Within section *Schiedea*, the nuclear BEST tree (Fig. 2) supports *S. spergulina*, then *S. hookeri* and *S. kealiae*, as sister to the remaining members of the section. This is not in conflict with the poorer resolution for this node on the concatenated nuclear species tree (Fig. 3). This position clarifies the direction of introgression that we can infer from cytonuclear conflict. The basal position of *S. spergulina* within section *Schiedea* suggests that it represents the sole surviving species on the older island of Kauaì, with the remainder of the radiation occurring as part of a migration onto the younger islands of Oáhu and Maui Nui.

Hard polytomies for remaining species in section *Schiedea* on both the nuclear BEST (Fig. 2) and the nuclear-plastid BEST (Fig. 4d) species trees, several paraphyletic species on the concatenated nuclear species tree (Fig. 3), conflicting resolution for these species based on methodology for the total-evidence species trees (Fig. 4a vs. Fig. 4b vs. Fig. 4c), taken together with weak resolution based on nrDNA and morphology (Wagner et al., 2005), suggests that their biologically realistic topology is very close to a hard polytomy. While the plastid lineage is more clearly resolved, it appears to be partly tracking geography rather than species lineages (Fig. 6). We emphasize that plastid conflict is located deep in the tree (*Alphaschiedea*), at the tips (e.g. *S. globosa* on Oáhu), and in intermediate positions (e.g. *S. diffusa*), requiring great caution in the interpretation of the chloroplast lineage. This finding is of great importance, as whole-plastome sequencing techniques currently being applied to other species, e.g. (Parks et al., 2009), will be alluring for their potential to resolve polytomies in groups such as *Schiedea*.

4.4. Original colonization and inter-island migrations

A molecular clock can be used to infer the age of a taxonomic group; reviewed by Sanderson et al. (2004). For Hawaiian archipelago radiations, an inference of the age of the root allows a ball-park estimate of how long ago the original colonization took place (Fleischer et al., 1998; Givnish et al., 2009; Jordan et al., 2003; Russo et al., 1995). The substantial divergence of the *Schiedea* chloroplast lineage from the species lineage (see Sections 3.3 and 4.3) needs to be considered in the interpretation of inferred node ages. Although there is clearly a geographic pattern, the *Schiedea* plastid lineage does not appear to be mainly tracking geography rather than species (Petit et al., 1993). Based on presumed maternal inheritance, chloroplasts are unlikely to accomplish inter-island migrations except within a species. Thus, their lineage can be used in *Schiedea* for a rough estimate of island progression.

Despite being rather well resolved at deeper nodes, our plastid phylogeny exhibits considerable rate heterogeneity among branches (see Section 2.6). This required the use of a relaxed log-normal clock that yields wider confidence intervals than expected from a clock-like tree (Supplementary Table S3). Although we were able to exclude root ages of 23 Ma and earlier based on a severe distortion of rates (see Section 2.6), root calibrations at 7.3 Ma and 4.7 Ma infer mean substitution rates (1.02 and 1.40×10^{-9} substitutions per site per year, respectively) that are near the range of $1.1–1.6 \times 10^{-9}$ reported for plastid loci across plants (Wolfe et al., 1987). Thus, we cannot use the criterion of unreasonable rates to exclude an original colonization of the *Schiedea* progenitor onto either Nihoa or Kauaì. Nonetheless, our BEAST results clarify several inter-island migration patterns.

All of our species trees (Figs. 2 and 4a–d) and the plastid tree (Fig. 1) place *S. verticillata*, a distinctive endemic from the leeward island of Nihoa, within the B clade. This topology is similar to a previous nrDNA and morphology tree (Wagner et al., 2005), except that our data infers that *S. verticillata*, rather than *S. attenuata*, is sister to remaining members of the B clade. Because we have confirmed the earlier branching of *Alphaschiedea*, *S. verticillata* is nested within the tree rather than representing the basal position that would be expected of a progenitor. We estimate the divergence of *S. verticillata* from the other members of the B clade at ca. 3 Ma (Figs. 6 and 7). This is more recent than the age of Kauaì and more recent than some other plastid lineages currently distributed on Kauaì. This suggests that *S. verticillata* represents a back-colonization from Kauaì to Nihoa rather than the maintenance of an ancestral lineage from the older islands.

Each of our analyses infers a migration of the ancestors of section *Alsinidendron* (*S. obovata* and *S. trinervis*) from Kauaì to Oáhu. Regardless of calibration scenario, *Alsinidendron* is inferred to have diverged from *Nothoschiedea* 1.7 Ma (Figs. 6 and 7). The plastid lineage shared by *S. spergulina* and *S. stellarioides* split ca. 0.5–0.9 Ma (Figs. 6 and 7). There is also a general consensus that the progenitor of the lineage leading to sections *Schiedea* and *Mononeura* migrated from Kauaì to Oáhu or from Kauaì to the islands of the Maui Nui complex ca. 2.5–2.1 Ma (Figs. 6 and 7). However, the pattern of inter-island dispersals within sections *Schiedea* and *Mononeura* to Oáhu and to the younger islands of Maui Nui is largely obscured by short branches and poorly resolved nodes that appear to be biologically realistic.

4.5. How old are the conflicting nodes?

Our BEAST results suggest that the TMRCA for the plastid lineage leading to the 19 species in section *Schiedea* and the MI clade (excepting *S. diffusa*) is at least 1.3 Ma (Fig. 7). We used this inferred age to judge whether these species would be expected to demonstrate incomplete lineage sorting for nuclear genes under a coalescent model. A coalescent time for 95% of loci to reach reciprocal monophyly can be estimated using a large population size ($N = 10{,}000$) and a reasonable generation time for *Schiedea* ($g = 5$; Wallace et al., 2009) to yield a conservative estimate (Maddison

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

and Knowles, 2006). Applying formulas of '9Ng' and '12Ng' (Hudson and Coyne, 2002), we calculate a span of 450,000–600,000 years until reciprocal monophyly is expected. Only a few *Schiedea* species pairs are inferred to have diverged within this timeframe (Fig. 7). It is also likely that *Schiedea* ancestral effective population sizes (for rare, insular species) were relatively small and we know that the typical census size for extant *Schiedea* species is much smaller than 10,000 (Wagner et al., 2005). Thus, the expected coalescent time may be considerably less than 450,000 years for *Schiedea*. Unless the original colonization was much more recent than our inference (Figs. 6 and 7) or *Schiedea* species lineages diverged much more recently than their plastid lineages, then our lack of reciprocal monophyly in nuclear gene trees within section *Schiedea* (Supplementary Figs. S1–S3) is unlikely to be due to incomplete lineage sorting alone. One explanation for this phenomenon is that reticulate evolution has played a major role in the radiation of section *Schiedea*. Additional, undetected, nuclear introgression may be masquerading as incomplete lineage sorting in this group.

We also considered whether the conflicting support inferred by parsimony and ML for different pairs of taxa in the total-evidence trees (Figs. 4a and b) is likely to be due to an anomaly zone created by short internal branches. Our conservatively calibrated plastid clock (Fig. 7) infers a TMRCA for these taxa pairs of 1.3 Ma (*S. salicaria*, *S. ligustrina*); 1.1 Ma (*S. adamantis*, *S. lydgatei* and *S. kealiae*, *S. ligustrina*); and ca. 0.65 Ma (*S. lydgatei*, *S. salicaria*). Using a population size of $N = 10{,}000$ and a generation time of $g = 5$, the total branch lengths in coalescent units ($t/2N$) can be roughly estimated at 6.5, 11, and 13, respectively. These branch lengths are much longer than ones expected to form an 'anomaly zone' (Kubatko and Degnan, 2007; their Fig. S1). Thus, anomalous trees that are distorted by short internal branches are unlikely to fully account for the conflict we observe in concatenated methods. Again, undetected introgression could offer an alternative explanation. Obviously, this coarse estimate ignores many complexities, including extinctions and short branches nested within the clade. We propose that these exercises are helpful by connecting empirical results with simulations in order to improve our understanding of real-world problems.

4.6. Evolution of breeding systems

An interest in reconstructing the pattern of breeding system shifts remains an important goal of a resolved phylogeny of *Schiedea*. Our placement of *S. apokremnos* sister to clade B (Fig. 2) differs from previous results (Wagner et al., 2005; Weller et al., 1998) and confirms that gynodioecy arose independently in this species. At minimum, one additional shift to dimorphic breeding systems occurred in section *Schiedea*. Unfortunately, the lack of resolution within this clade does not support a formal reconstruction of breeding system character states. In fact, incongruence among our four independent genetic partitions should suggest caution in using a single genetic partition to infer character states in section *Schiedea*. We cannot reject either possibility: (i) hermaphroditism as the plesiomorphic state for the section and an unknown number of shifts to dimorphy; or (ii) dimorphy as the plesiomorphic state for the section and up to three reversals to hermaphroditism (*S. lydgatei*, *S. menziesii*, and *S. hookeri*). The nuclear BEST tree (Fig. 2) places the hermaphroditic *S. hookeri* outside the rake, but this is not supported by our nuclear concatenated tree (Fig. 3). None of our species trees places either of the other two hermaphroditic species (*S. lydgatei* and *S. menziesii*) on an early branch within the section. In fact, the dioecious *S. spergulina* is basal in the nuclear BEST tree (Fig. 2). A likelihood ratio test constraining the three hermaphroditic species to monophyly (and hence representing one reversal) is not statistically different than an unconstrained tree (data not shown) suggesting that such a scenario is not unwarranted. Although the pattern of interspecific breeding system shifts in section *Schiedea* remains complex, intraspecific studies of gynodioecious *Schiedea* species are a useful system in which to study the transition from hermaphroditism toward dioecy (Campbell et al., 2010; Rankin et al., 2002).

5. Conclusions

Although all 32 extant *Schiedea* species are monophyletic, our trees reveal patterns of introgression suggesting that interspecific hybridization has exerted a strong influence, despite being apparently infrequent. An example of reticulate evolution occurs deep in the lineage of this genus, and there are several more examples toward the tips. The use of cytonuclear conflict among individual gene trees was not reliable for identifying divergent plastid lineages to remove from this data set. In particular, the paraphyletic nature of *S. kauaiensis* was a misleading clue. A comparison of the plastid tree with better-resolved nuclear species trees indicates that the lateral transfer of a plastid lineage likely occurred in the opposite direction, i.e. that the basally-branching section *Alphaschiedea* acquired a chloroplast lineage from the more recent section *Mononeura*. Our results are equivocal for nuclear introgression. Molecular clock calibrations suggest an unexpectedly ancient age (ca. 2–2.5 Ma) for the chloroplast lineage of the crown radiation of sections *Mononeura* and *Schiedea*. Because this timeframe exceeds the expected coalescence time for species with small effective population sizes, the lack of resolution among section *Schiedea* species has likely been driven, at least partially, by additional chloroplast transfers.

In the presence of this level of reticulate evolution, the plastid lineage, our nuclear-plastid BEST tree, and our total-evidence approach are positively misleading in their inferred species trees, despite our use of a 'compromise' data set. Further, we obtained stronger sectional resolution with a nuclear BEST tree compared with a concatenated species tree method. Whether other nodes on the nuclear concatenated tree that lack support on the nuclear BEST tree will withstand remains to be seen. In any case, it appears that additional nuclear evidence, and perhaps the application of population-structure models, will be needed to tease apart the pattern of breeding system evolution within section *Schiedea*.

Acknowledgments

This research was supported by a Grant from the National Science Foundation (DEB-0445410). The authors thank the National Tropical Botanical Gardens McBryde Senior Research program for support to MN; Douglas Jennewein and Joe Reynoldson for computing support; Tim Flynn, Joel Lau, John Obata, Steve Perlman, Wayne Takeuchi, and Ken Wood for field collections; Tina Clausen, Jeff Engeman, Bridget Jacobs, Calvin Krogman, and Robynn Shannon for laboratory and herbarium work; L. Lacey Knowles, Laura Kubatko and the participants in the Estimating Species Trees workshop for inspiration; Adam Bazinet and Liang Liu for helpful suggestions on software implementation.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.04.001.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19, 716–723.

Belfiore, N.M., Liu, L., Moritz, C., 2008. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). Syst. Biol. 57, 294–310.
Bossu, C.M., Near, T.J., 2009. Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (Percidae: *Etheostoma*). Syst. Biol. 58, 114–129.
Brown, R.P., Yang, Z., 2010. Bayesian dating of shallow phylogenies with a relaxed clock. Syst. Biol. 59, 119–131.
Brumfield, R.T., Liu, L., Lum, D.E., Edwards, S.V., 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded Manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data. Syst. Biol. 57, 719–731.
Campbell, D.R., Weller, S.G., Sakai, A.K., Culley, T.M., Dang, P.N., Dunbar-Wallis, A.K., 2010. Genetic variation and covariation in floral allocation of two species of *Schiedea* with contrasting levels of sexual dimorphism. Evolution 65, 757–770.
Caraway, V., Carr, G.D., Morden, C., 2001. Assessment of hybridization and introgression in lava-colonizing Hawaiian *Dubautia* (Asteraceae: Madiinae) using RAPD markers. Am. J. Bot. 88, 1688–1694.
Carlquist, S., 1974. Island Biology. Columbia University Press, New York, NY, USA.
Carstens, B.C., Knowles, L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Syst. Biol. 56, 400–411.
Chan, K.M., Levin, S.A., 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59, 720–729.
Clague, D.A., 1996. The growth and subsidence of the Hawaiian-Emperor volcanic chain. In: Keast, A., Miller, S.E. (Eds.), The Origin and Evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: Patterns and Processes. SPB Academic Publishing, Amsterdam, Netherlands, pp. 35–50.
Clark, A., 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. 7, 111–122.
Culley, T.M., Dunbar-Wallis, A.K., Sakai, A.K., Weller, S.G., Mishio, M., Campbell, D.R., Herzenach, M., 2006. Genetic variation of ecophysiological traits in two gynodioecious species of *Schiedea* (Caryophyllaceae). New Phytol. 169, 589–601.
Cummings, M., Neel, M., Shaw, K., Bazinet, A., 2010. Genealogical Sorting Index (Web-based Software Program). <http://gsi.umiacs.umd.edu/> (accessed 11.06.10).
Cummings, M.P., Neel, M.C., Shaw, K.L., 2008. A genealogical approach to quantifying lineage divergence. Evolution 62, 2411–2422.
Currat, M., Ruedi, M., Petit, R.J., Excoffier, L., 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62, 1908–1920.
DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. Syst. Biol. 49, 171–179.
Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics 2, 762–768.
Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–131.
Doolittle, W.F., Bapteste, E., 2007. Pattern pluralism and the Tree of Life hypothesis. Proc. Natl. Acad. Sci. USA 104, 2043–2049.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214–221.
Dunbar-Co, S., Wieczorek, A.M., Morden, C.W., 2008. Molecular phylogeny and adaptive radiation of the endemic Hawaiian *Plantago* species (Plantaginaceae). Am. J. Bot. 95, 1177–1188.
Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.
Edwards, S.V., Beerli, P., 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54, 1839–1854.
Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.
Emshwiller, E., Doyle, J.J., 1999. Chloroplast-expressed glutamine synthetase (*ncpGS*): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). Mol. Phylogen. Evol. 12, 310–319.
Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Assoc., Inc., Sunderland, MA, USA.
Fleischer, R.C., McIntosh, C.E., Tarr, C.L., 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K–Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. Mol. Ecol. 7, 533–545.
Frajman, B., Eggens, F., Oxelman, B., 2009. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae) – a multigene phylogenetic approach with relative dating. Syst. Biol. 58, 328–345.
Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Evol. Syst. 34, 397–423.
Funk, V.A., Wagner, W.L., 1995. Biogeographic patterns in the Hawaiian Islands In: Wagner, W.L., Funk, V.A. (Eds.), Hawaiian Biogeography: Evolution on A Hot Spot Archipelago. Smithsonian Institution Press, Washington, DC, USA.
Gaskin, J.F., Schaal, B.A., 2002. Hybrid *Tamarix* widespread in US invasion and undetected in native Asian range. Proc. Natl. Acad. Sci. USA 99, 11256–11259.
Givnish, T.J., Sytsma, K.J., Smith, J.F., Hahn, W.J., 1995. Molecular evolution, adaptive radiation and geographic speciation in *Cyanea* (Campanulaceae, Lobelioideae). In: Wagner, W.L., Funk, V.A. (Eds.), Hawaiian Biogeography: Evolution in a Hotspot Archipelago. Smithsonian Institution Press, Washington, DC, USA, pp. 288–336.
Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R., Sytsma, K.J., 2009. Origin, adaptive radiation, and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). Proc. R. Soc. Lond. B 276, 407–416.
Golonka, A.M., Sakai, A.K., Weller, S.G., 2005. Wind pollination, sexual dimorphism, and changes in floral traits of *Schiedea* (Caryophyllaceae). Am. J. Bot. 92, 1492–1502.
Goodman, M., Czelusniak, J., Moore, G.W., Romero-Herrera, A.E., Matsuda, G., 1979. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from Globin sequences. Syst. Zool. 28, 132–163.
Grant, V., 1981. Plant Speciation. Columbia Univ. Press, New York, NY, USA.
Havran, J.C., Sytsma, K.J., Ballard Jr., H.E., 2009. Evolutionary relationships, interisland biogeography, and molecular evolution in the Hawaiian violets (*Viola*: Violaceae). Am. J. Bot. 96, 2087–2099.
Heled, J., Drummond, A.J., 2009. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.
Howarth, D.G., Baum, D.A., 2002. Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. Mol. Phylogen. Evol. 23, 525–528.
Howarth, D.G., Baum, D.A., 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaveola* (Goodeniaceae) in the Hawaiian Islands. Evolution 59, 948–961.
Huang, H.T., Knowles, L.L., 2009. What is the danger of the anomaly zone for empirical phylogenetics? Syst. Biol. 58, 527–536.
Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. Evolution 56, 1557–1565.
Johnson, L.A., Soltis, D.E., 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae *s. str*. Syst. Bot. 19, 143–156.
Jordan, S., Simon, C., Polhemus, D., 2003. Molecular systematics and adaptive radiation of Hawaii's endemic Damselfly genus *Megalagrion* (Odonata: Coenagrionidae). Syst. Biol. 52, 89–109.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.
Leache, A.D., 2009. Species tree discordance traces to phylogeographic clade boundaries in North American Fence Lizards (*Sceloporus*). Syst. Biol. 58, 547–559.
Lee, C., Wen, J., 2004. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. Mol. Phylogen. Evol. 31, 894–903.
Lindqvist, C., Motley, T.J., Jeffrey, J.J., Albert, V.A., 2003. Cladogenesis and reticulation in the Hawaiian endemic mints (Lamiaceae). Cladistics 19, 480–495.
Linnen, C.R., Farrell, B.D., 2008. Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). Syst. Biol. 57, 876–890.
Liu, L., 2008. BEST: Bayesian estimation of species under the coalescent model. Bioinformatics 24, 2542–2543.
Liu, L., Edwards, S.V., 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58, 452–460.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.
Liu, L., Pearl, D.K., Brumfield, R.T., Edwards, S.V., 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62, 2080–2091.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.
Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.
Moore, G., Devos, K., Wang, Z., Gale, M., 1995. Grasses, line up and form a circle. Curr. Biol. 5, 737.
Müller, K., 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinf. 4, 65–69.
Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.
Norman, J.K., Sakai, A.K., Weller, S.G., Dawson, T.E., 1995. Inbreeding depression in morphological and physiological traits of *Schiedea lydgatei* (Caryophyllaceae) in two environments. Evolution 49, 297–306.
Norman, J.K., Weller, S.G., Sakai, A.K., 1997. Pollination biology and outcrossing rates in hermaphroditic *Schiedea lydgatei* (Caryophyllaceae). Am. J. Bot. 84, 641–648.
Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7, 84–100.
Petit, R.J., Kremer, A., Wagner, D.B., 1993. Geographic structure of chloroplast DNA polymorphisms in European oaks. Theor. Appl. Gen. 87, 122–128.
Pirie, M.D., Humphreys, A.M., Barker, N.P., Linder, H.P., 2009. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). Syst. Biol. 58, 612–628.
Popp, M., Oxelman, B., 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Mol. Phylogen. Evol. 20, 474–481.
Popp, M., Erixon, P., Eggens, F., Oxelman, B., 2005. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001

from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. Syst. Bot. 30, 302–313.
Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Price, J.P., Clague, D.A., 2002. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. Proc. R. Soc. Lond. B 269, 2429–2435.
Rambaut, A., 2010a. FigTree (Software Program). <http://tree.bio.ed.ac.uk/software/> (accessed 11.06.10).
Rambaut, A., 2010b. Se–Al, Sequence Alignment Editor (Software Program). <http://iubio.bio.indiana.edu/soft/iubionew/molbio/dna/analysis/Pist/main.html/> (accessed 11.06.10).
Rankin, A.E., Weller, S.G., Sakai, A.K., 2002. Mating system instability in *Schiedea menziesii* (Caryophyllaceae). Evolution 56, 1574–1585.
Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants 5, 65–84.
Rieseberg, L.H., Whitton, J., Linder, C.R., 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Bot. Neerl. 45, 243–262.
Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.
Russo, C.A.M., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12, 391–404.
Sakai, A.K., Weller, S.G., 1991. Ecological aspects of sex expression in subdioecious *Schiedea globosa* (Caryophyllaceae). Am. J. Bot. 78, 1280–1288.
Sakai, A.K., Karoly, K., Weller, S.G., 1989. Inbreeding depression in *Schiedea globosa* and *S. salicaria* (Caryophyllaceae), subdioecious and gynodioecious Hawaiian species. Am. J. Bot. 76, 437–444.
Sakai, A.K., Weller, S.G., Chen, M.-L., Chou, S.-Y., Tasanont, C., 1997. Evolution of gynodioecy and maintenance of females: the role of inbreeding depression, outcrossing rates, and resource allocation in *Schiedea adamantis* (Caryophyllaceae). Evolution 51, 724–736.
Sakai, A.K., Weller, S.G., Wagner, W.L., Nepokroeff, M., Culley, T.M., 2006. Adaptive radiation and evolution of breeding systems in *Schiedea* (Caryophyllaceae), an endemic Hawaiian genus. Ann. Missouri Bot. Gard. 93, 49–63.
Sakai, A.K., Weller, S.G., Culley, T.M., Campbell, D.R., Dunbar-Wallis, A.K., Andres, A., 2008. Sexual dimorphism and the genetic potential for evolution of sex allocation in the gynodioecious plant, *Schiedea salicaria*. J. Evol. Biol. 21, 18–29.
Sanderson, M., Thorne, J., Wikstrom, N., Bremer, K., 2004. Molecular evidence on plant divergence times. Am. J. Bot. 91, 1656–1665.
Sang, T., Zhong, Y., 2000. Testing hybridization hypotheses based on incongruent gene trees. Syst. Biol. 49, 422–434.
Schneider, H., Ranker, T.A., Russell, S.J., Cranfill, R., Geiger, J.M.O., Aguraiuja, R., Wood, K.R., Grundmann, M., Kloberdanz, K., Vogel, J.C., 2005. Origin of the endemic fern genus *Diellia* coincides with the renewal of Hawaiian terrestrial life in the Miocene. Proc. R. Soc. Lond. B 272, 455–460.
Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207.
Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. USA 99, 16122–16127.
Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166.
Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288.
Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.
Soltis, P.S., Soltis, D.E., Weller, S.G., Sakai, A.K., Wagner, W.L., 1996. Molecular phylogenetic analysis of the Hawaiian endemics *Schiedea* and *Alsinidendron* (Caryophyllaceae). Syst. Bot. 21, 365–379.
Spinks, P.Q., Shaffer, H.B., 2009. Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. Syst. Biol. 58, 1–20.
Stamatakis, A., 2006. RAxML-V1-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stebbins, G.L., 1950. Variation and Evolution in Plants. Columbia Univ. Press, New York, NY.
Swofford, D., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0 b10. Sinauer Assoc., Inc., Sunderland, MA, USA.
Taylor, E.B., McPhail, J.D., 2008. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. Biol. J. Linn. Soc. 66, 271–291.
Thiers, B., 2010. *Index Herbariorum* (Online Database). <http://sweetgum.nybg.org/ih/> (accessed 11.06.10).
Wagner, W.L., Weller, S.G., Sakai, A.K., 1995. Phylogeny and biogeography in *Schiedea* and *Alsinidendron* (Caryophyllaceae). In: Wagner, W.L., Funk, V.A. (Eds.), Hawaiian Biogeography. Smithsonian Institution Press, Washington, DC, pp. 221–258.
Wagner, W.L., Herbst, D.R., Sohmer, S., 1999. Manual of the Flowering Plants of Hawai'i. Bishop Museum Press, Honolulu, HI, USA.
Wagner, W.L., Weller, S.G., Sakai, A.K., 2005. Monograph of *Schiedea* (Caryophyllaceae–Alsinoideae). Syst. Bot. Monographs 72, 169.
Wallace, L.E., Weller, S.G., Wagner, W.L., Sakai, A.K., Nepokroeff, M., 2009. Phylogeographic patterns and demographic history of *Schiedea globosa* (Caryophyllaceae) on the Hawaiian Islands. Am. J. Bot. 96, 958–967.
Wares, J.P., Pankey, M.S., Pitombo, F., Gómez Daglio, L., Achituv, Y., 2009. A "shallow phylogeny" of shallow barnacles (*Chthamalus*). PLoS One 4, 1–10.
Weller, S.G., Sakai, A.K., 1991. The genetic basis of male sterility in *Schiedea* (Caryophyllaceae), an endemic Hawaiian genus. Heredity 67, 265–273.
Weller, S.G., Sakai, A.K., 1999. Using phylogenetic approaches for the analysis of plant breeding system evolution. Ann. Rev. Ecol. Syst. 30, 167–199.
Weller, S.G., Sakai, A.K., 2005. Selfing and resource allocation in *Schiedea salicaria* (Caryophyllaceae), a gynodioecious species. J. Evol. Biol. 18, 301–308.
Weller, S.G., Sakai, A.K., Wagner, W.L., Herbst, D.R., 1990. Evolution of dioecy in *Schiedea* (Caryophyllaceae: Alsinoideae) in the Hawaiian Islands: biogeographical and ecological factors. Syst. Bot. 15, 266–276.
Weller, S.G., Wagner, W.L., Sakai, A.K., 1995. A phylogenetic analysis of *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae): implications for the evolution of breeding systems. Syst. Bot. 20, 315–337.
Weller, S.G., Sakai, A.K., Straub, C., 1996. Allozyme diversity and genetic identity in *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae) in the Hawaiian Islands. Evolution 50, 23–34.
Weller, S.G., Sakai, A.K., Rankin, A.E., Golonka, A.M., Kutcher, B., Ashby, K.E., 1998. Dioecy and the evolution of pollination systems in *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae) in the Hawaiian Islands. Am. J. Bot. 85, 1377–1388.
Weller, S.G., Sakai, A.K., Wagner, W.L., 2001. Artificial and natural hybridization in *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae): the importance of phylogeny, genetic divergence, breeding system, and population size. Syst. Bot. 26, 571–584.
Weller, S.G., Sakai, A.K., Culley, T.M., Campbell, D.R., Dunbar-Wallis, A.K., 2006. Predicting the pathway to wind pollination: heritabilities and genetic correlations of inflorescence traits associated with wind pollination in *Schiedea salicaria* (Caryophyllaceae). J. Evol. Biol. 19, 331–342.
Weller, S.G., Sakai, A.K., Culley, T.M., Campbell, D.R., Ngo, P., Dunbar-Wallis, A.K., 2007. Sexually dimorphic inflorescence traits in a wind-pollinated species: heritabilities and genetic correlations in *Schiedea adamantis* (Caryophyllaceae). Am. J. Bot. 94, 1716–1725.
Willyard, A., Cronn, R., Liston, A., 2009. Reticulate evolution among the ponderosa pines. Mol. Phylogen. Evol. 52, 498–511.
Wolfe, K., Li, W.-H., Sharp, P., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Please cite this article in press as: Willyard, A., et al. Estimating the species tree for Hawaiian *Schiedea* (Caryophyllaceae) from multiple loci in the presence of reticulate evolution. Mol. Phylogenet. Evol. (2011), doi:10.1016/j.ympev.2011.04.001