

Plastid genomes reveal recurrent formation of allopolyploid *Fragaria*

Markus S. Dillenberger¹, Na Wei², Jacob A. Tennessen³, Tia-Lynn Ashman², and Aaron Liston^{1,4}

Manuscript received 12 October 2017; revision accepted 7 February 2018.

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

² Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA

³ Department of Integrative Biology, Oregon State University, Corvallis, OR, USA

⁴ Author for correspondence (e-mail: listona@science.oregonstate.edu)

Citation: Dillenberger, M. S., N. Wei, J. A. Tennessen, T.-L. Ashman, and A. Liston. 2018. Plastid genomes reveal recurrent formation of allopolyploid *Fragaria*. *American Journal of Botany* 105(5): 1–13.

doi:10.1002/ajb2.1085

PREMISE OF THE STUDY: Recurrent formation of polyploid taxa is a common observation in many plant groups. Haploid, cytoplasmic genomes like the plastid genome can be used to overcome the problem of homeologous genes and recombination in polyploid taxa. *Fragaria* (Rosaceae) contains several octo- and decaploid species. We use plastome sequences to infer the plastid ancestry of these taxa with special focus on the decaploid *Fragaria cascadiensis*.

METHODS: We used genome skimming of 96 polyploid *Fragaria* samples on a single Illumina HiSeq 3000 lane to obtain whole plastome sequences. These sequences were used for phylogenetic reconstructions and dating analyses. Ploidy of all samples was inferred with flow cytometry, and plastid inheritance was examined in a controlled cross of *F. cascadiensis*.

KEY RESULTS: The plastid genome phylogeny shows that only the octoploid *F. chiloensis* is monophyletic, all other polyploid taxa were supported to be para- or polyphyletic. The decaploid *Fragaria cascadiensis* has biparental plastid inheritance and four different plastid donors. Diversification of the *F. cascadiensis* clades occurred in the last 230,000 years. The southern part of its distribution range harbors considerably higher genetic diversity, suggestive of a potential refugium.

CONCLUSIONS: *Fragaria cascadiensis* had at least four independent origins from parents with different plastomes. In contrast, para- and polyphyletic taxa of the octoploid *Fragaria* species are best explained by incomplete lineage sorting and/or hybridization. Biogeographic patterns in *F. cascadiensis* are probably a result of range shift during the last glacial maximum.

KEY WORDS allopolyploid; biparental plastid inheritance; *Fragaria cascadiensis*; genome skimming; independent origins; plastome; refugia; Rosaceae.

Allopolyploid taxa are the result of hybridization between different parental taxa in combination with a whole genome duplication (Jenkins and Rees, 1991). Genetic incompatibility and meiotic pairing problems can lead to reduced fertility in homoploid hybrids, but polyploidization can overcome these difficulties (Mallet, 2007; Rieseberg and Willis, 2007; Paun et al., 2009). Although an allopolyploid hybrid species theoretically may be the result of a single hybridization event, it is likely that hybridization has occurred repeatedly in the contact zone of the parental species (Soltis and Soltis, 1999). In fact, a major advance in the research of polyploid species was the recognition that most allopolyploid species have multiple origins (reviewed in Soltis and Soltis, 1999; Soltis et al., 2014). Furthermore, the documentation of ubiquitous ancient whole genome duplications has transformed our understanding of angiosperm evolutionary history (Fawcett et al., 2009; Wendel, 2015). In the past years, increasing numbers of studies used

genomic approaches to investigate the question of recurrent formation of polyploids in plants (e.g., Folk et al., 2017; Sherman-Broyles et al., 2017). Groups in which recurrent hybrid formation has been detected include tetraploid *Tragopogon* L. (Soltis et al., 2004), diploid and tetraploid *Heuchera* L. (Segraves et al., 1999), diploid *Helianthus* L. (Schwarzbach and Rieseberg, 2002), and tetraploid *Glycine* Willd. (Sherman-Broyles et al., 2017). Many of these studied polyploid species with recurrent formation are recently formed (e.g., some *Tragopogon* hybrids; Soltis et al., 2004). Evolutionary mechanisms like selection and/or genetic drift can obscure the ancestry of allopolyploid hybrids over time. Newly formed hybrids are labile and likely have higher extinction rates. Thus, genetic diversity acquired by recurrent hybridization might get lost shortly after hybrid formation. Older hybrid lineages have probably lost some degree of their initial genetic diversity because of these effects, and the number of recurrent formations might be underestimated.

Independent origins of polyploid taxa can be investigated by elucidating the origin of the plastid genome (e.g., Schwarzbach and Rieseberg, 2002). If a polyploid hybrid is the result of recurrent formation with parents having different plastid haplotypes, phylogenetic analyses of the plastome should show multiple clades of the hybrid taxon each most closely related to their plastid donor. A single origin of a hybrid taxon can be assumed if only one clade is found in a phylogeny. Early studies using a set of standard plastid markers often suffered from a lack of resolution (e.g., Brochmann et al., 1992; Sang et al., 1997). Complete plastomes can overcome the limitation of insufficient sequence variation (Parks et al., 2009; Bernhardt et al., 2017; Spooner et al., 2017). There are few studies using whole plastome phylogenies based on range-wide sampling with multiple populations (Huang et al., 2014; Pham et al., 2017). With our approach, we were able to elucidate the evolutionary history from the individual to species level.

Plastids and mitochondria are often assumed to be inherited from the maternal parent (Mogensen, 1996; Zhang et al., 2003), but this assumption often remains untested in population level studies. Corriveau and Coleman (1988) and Zhang et al. (2003) investigated large numbers of species for the potential of male inheritance of cytoplasm. Both studies found that approximately 80% of the species have maternal cytoplasmic inheritance, and about 20% showed the potential for biparental inheritance. Paternal inheritance of plastids was repeatedly found for a variety of taxa (e.g., Szmidi et al., 1987; Neale and Sederoff, 1989; Reboud and Zeyl, 1994; Svab and Maliga, 2007; Ellis et al., 2008). Biparental inheritance does not necessarily mean that the offspring have two types of plastids. Matsushima et al. (2008) showed in the case of *Medicago* L. that seedlings with both parental types of plastids had only one type of plastids in adult tissue.

Here we utilize genome skimming (Straub et al., 2014) to determine the sequence of plastid genomes of the North American polyploid *Fragaria* L. species. This clade comprises three octoploid and one decaploid species. Previous studies showed that these species form a clade in plastid phylogenies (Njuguna et al., 2013; Govindarajulu et al., 2015), but sampling of taxa and populations was limited and incomplete. Here we substantially increase sampling compared to previous studies, and particularly focus on a decaploid species, *F. cascadenensis* (see Table 1 for authors of frequently used names) which was only recently identified and thus not previously put in a phylogenetic context. As in most flowering plants,

the plastid is assumed to be maternally inherited in *Fragaria* (Davis et al., 2010; Govindarajulu et al., 2013), but so far it was only investigated for diploid species of *Fragaria* (Davis et al., 2010). There are no previous studies investigating plastid inheritance in *F. cascadenensis* or other North American polyploid *Fragaria* species.

The North American polyploids comprise four species. (1) *Fragaria cascadenensis*, the only decaploid species in North America, was described in 2012 and is an endemic of the Oregon Cascade Mountains (Hummer, 2012). (2) *Fragaria chiloensis* is octoploid and has three subspecies in North and South America and on Hawaii (Liston et al., 2014). (3) *Fragaria virginiana* is also octoploid and has four subspecies in North America. (4) *Fragaria ×ananassa* is the octoploid hybrid of *F. chiloensis* and *F. virginiana* (Liston et al., 2014) and has a naturally occurring subspecies in northwestern North America, which can be distinguished from the cultivated strawberry. The ancestry of the North American octoploids has recently been clarified. Tennessen et al. (2014) showed that one of the four subgenomes is derived from *F. vesca* subsp. *bracteata*—one from *F. iinumae* and two from an extinct ancestor closely related to *F. iinumae*. For the decaploid *F. cascadenensis*, Hummer (2012) hypothesized that the additional subgenome belongs to *F. vesca* subsp. *bracteata* based on geography and intermediate combination of morphological traits, but phylogenetic analysis of linkage-mapped chromosomes from targeted sequences (Wei et al., 2017) showed instead that it is a third *F. iinumae*-like genome. The previous study by Wei et al. (2017) focused on the nuclear genome of the decaploid and thus did not test the origin of the plastid genome (i.e., the octoploid vs. diploid progenitor). Husband et al. (2002) showed that 2× pollen is more successful in fertilization than 1× pollen for *Chamerion angustifolium* (L.) Holub, indicating that polyploids may be more successful paternal parents, but knowledge is sparse and general conclusions might be premature (Kreiner et al., 2017).

The plastid donor of the octoploids is *F. vesca* subsp. *bracteata* (Njuguna et al., 2013). Although it is known that *F. chiloensis*, *F. virginiana* and their hybrid *F. ×ananassa* are closely related and have the same set of subgenomes (Tennessen et al., 2014), previous phylogenies lack conclusive results regarding monophyly and status of taxa due to limited sampling and evidence based on a small number of crosses.

In addition to a representative sampling of diploid and octoploid species and a dense sampling of *F. cascadenensis*, we included the only other known decaploid in the genus, *F. iturupensis*. This

TABLE 1. Taxon sampling in *Fragaria* and the outgroup used in this study.

Taxon	Ploidy	# of populations	# of samples	Newly sequenced
<i>Dasiphora floribunda</i> (Pursh) Raf.	2x	1	1	0
<i>Fragaria xananassa</i> (Weston) Duchesne ex Rozier subsp. <i>ananassa</i>	8x	1	1	0
<i>F. xananassa</i> (Weston) Duchesne ex Rozier subsp. <i>cuneifolia</i> (Nutt. ex Howell) Staudt	8x	2	4	1
<i>F. cascadenensis</i> K.E. Hummer	10x	16	81	79
Nonaploid <i>Fragaria</i>	9x	3	8	8
<i>F. chiloensis</i> (L.) Mill. subsp. <i>chiloensis</i>	8x	1	1	0
<i>F. chiloensis</i> (L.) Mill. subsp. <i>pacifica</i> Staudt (incl. subsp. <i>lucida</i> (E. Vilm. ex J. Gay) Staudt)	8x	5	20	0
<i>F. iinumae</i> Makino	2x	1	1	0
<i>F. iturupensis</i> Staudt	10x	1	1	1
<i>F. mandshurica</i> Staudt	2x	1	1	0
<i>F. vesca</i> L. subsp. <i>bracteata</i> (A. Heller) Staudt	2x	2	5	0
<i>F. vesca</i> L. subsp. <i>vesca</i>	2x	1	1	0
<i>F. virginiana</i> Mill. subsp. <i>glauca</i> (S. Watson) Staudt	8x	3	3	0
<i>F. virginiana</i> Mill. subsp. <i>platypetala</i> (Rydb.) Staudt	8x	11	22	5
<i>F. virginiana</i> Mill. subsp. <i>virginiana</i>	8x	9	18	0

species is restricted to Iturup, a 3100 km² island in the southern Kuril Archipelago, ~125 km northeast of Hokkaido, Japan. The species was originally described as an octoploid (Staudt and Olbricht, 2008), but the plant used in the present study was confirmed to be a decaploid (Hummer et al., 2009). Phylogenetic analyses indicate that *F. iturupensis* shares the same plastid ancestry as the octoploid species (Njuguna et al., 2013) and suggest a hybrid origin between an octoploid and *F. iinumae* (Kamneva et al., 2017).

With extensive population-level sampling of *F. cascadenis* and well-represented sampling of the North American octoploid taxa, we were able to address the following questions: (1) Do the North American polyploid taxa, especially *F. cascadenis*, have single

origins? (2) Is the plastid strictly maternally inherited in *F. cascadenis*? (3) What is the identity (taxon and/or ploidy) of the plastid ancestry of *F. cascadenis*? (4) Is there a phylogeographic pattern in *F. cascadenis*?

MATERIALS AND METHODS

Study species and sampling strategy

Fragaria cascadenis is found in montane habitats between 900 and 1800 m in the Cascade Mountains of Oregon, USA (Fig. 1).

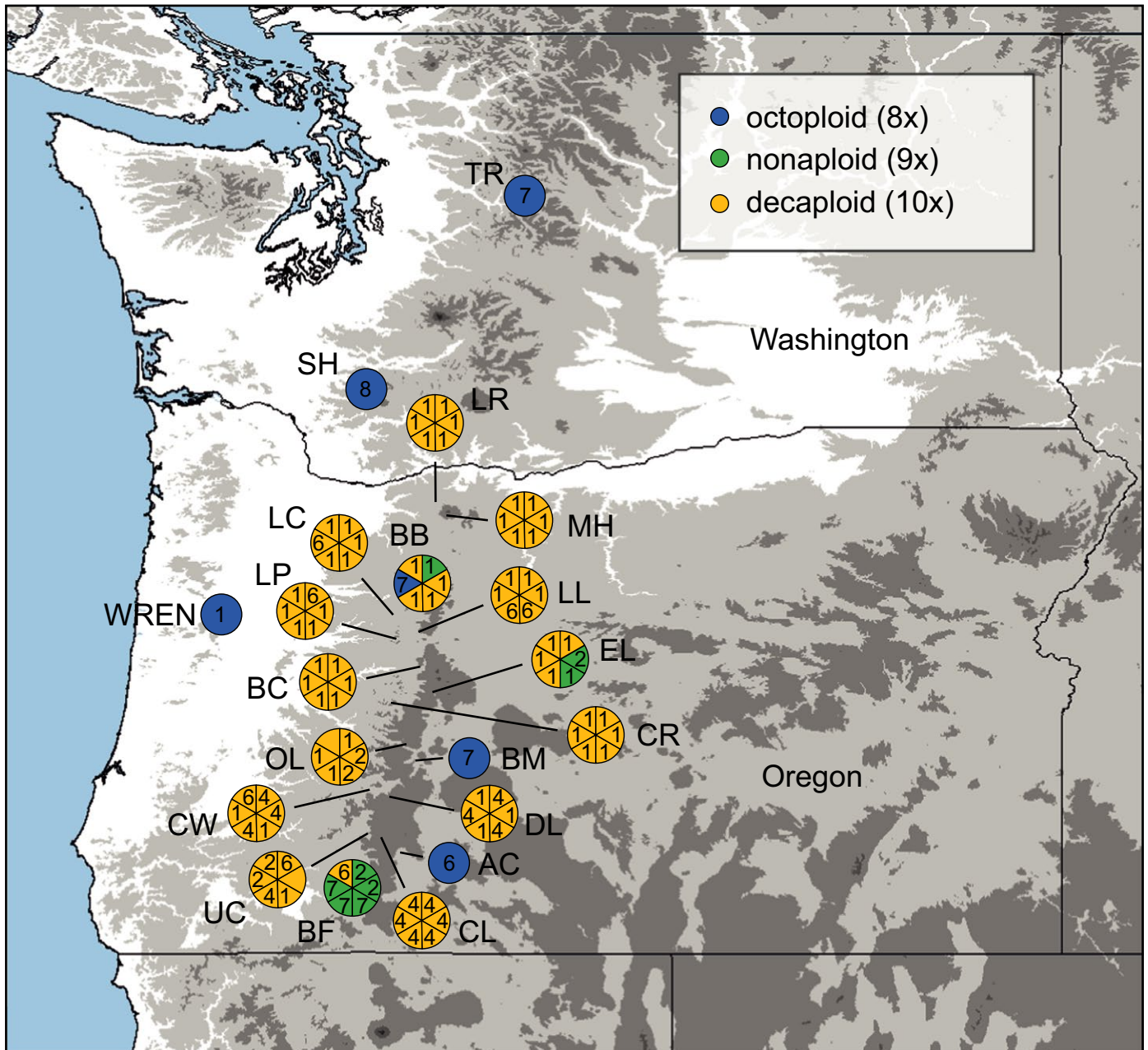


FIGURE 1. Distribution map of the sampled *F. cascadenis* populations in Table 1. The pie chart's colors indicate ploidy level. Numbers indicate clade affiliation according to Fig. 2 and Appendix S2. Population abbreviations are according to Table 1 and Appendix S1. All populations used in this study shown in Appendix S5.

The species was described by Hummer (2012), who distinguished *F. cascadiensis* by ploidy from sympatric octoploid *F. virginiana* subsp. *platypetala* and diploid *F. vesca* subsp. *bracteata*, but also found morphological differences between these taxa in the Oregon Cascades. *Fragaria cascadiensis* has hairs on the upper surface of the leaves, and achenes that are frequently comma-shaped, *F. virginiana* subsp. *platypetala* of the Oregon Cascades has no hairs on the upper leaf surface, and achenes are dome-shaped. Differences between *F. cascadiensis* and *F. vesca* subsp. *bracteata* are related to leaf color and leaf margins. *Fragaria cascadiensis* has slightly bluish green leaves and the distal tooth of terminal leaflets is smaller than surrounding teeth, while *F. vesca* subsp. *bracteata* has Kelly green leaves and the terminal tooth is usually longer than surrounding teeth (Hummer, 2012).

We sampled 15 populations of *F. cascadiensis* encompassing its entire ~330 km geographic north-south distribution (Table 2; Fig. 1). Each population is represented by six individuals. Field identification to differentiate *F. cascadiensis* from octoploid *F. virginiana* was based on the presence of hairs on the upper leaflet surfaces. Additionally, we included four samples of *F. virginiana* subsp. *platypetala*, one sample of *F. xananassa* subsp. *cuneifolia*, and one sample of *F. iturupensis* (Table 2, Appendix S1, see Supplemental Data with this article). To complement our data set, we included whole genome sequence data from an additional 61 samples of octoploid *Fragaria* species (Table 1; Tennessen et al., 2017), two additional samples of *F. cascadiensis* (Wei et al., 2017), samples of *F. xananassa* subsp. *ananassa* and *F. chiloensis* subsp. *chiloensis* from the NCBI Sequence Read Archive, and previously published plastome sequences from nine diploid species, including the two extant progenitors of the octoploid species (*F. vesca* subsp. *bracteata*, *F. iinumae*; Appendix S1).

Flow Cytometry

Ploidy level was determined on all samples collected for this study by flow cytometry (Appendix S1). We sampled runners or petioles from each plant. We used an internal octoploid standard (*F. virginiana* or *F. chiloensis*) for calibration. Samples (~0.5–0.7 cm of fresh stolon or petiole) were prepared according to Straub et al. (2011) and stained with CyStain UV Precise P 05-5002 (New England Biolabs Inc., Ipswich, Massachusetts, USA). The samples were cut in 150 µL nuclei extraction buffer (NEB), stained with 800 µL diluted stain (1:1 Staining Buffer : ddH₂O) and analyzed on a Partec PA (Sysmex Partec GmbH, Görlitz, Germany).

DNA extraction

Total genomic DNA was isolated from fresh/frozen leaf material collected in the field using the FastDNAKit (Qbiogene, Inc., California, USA) following the manufacturer's protocol with one modification. We added polyvinylpyrrolidone (PVP) and β-mercaptoethanol during cell lysis to a final concentration of 9.5 mg/L and 0.14 M, respectively. DNA was purified using ethanol precipitation (Sambrook and Russell, 2001). Each extraction was evaluated first on a 1% agarose gel and then quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, California, USA).

Sequencing library preparation

We used a Bioruptor Pico (Diagenode SA, Seraing (Ougrée), Belgium) to shear 100 µL of the 96 samples, containing between

TABLE 2. Newly sequenced populations of *Fragaria* in this study. Taxa, number of samples, and additional information for each population can be seen in Fig. 1 and Appendix S1.

Abbreviation	Population name	Population location	Population specimens
AC	Annie Creek	42°45'52.6"N 122°3'42.0"W	Dillenberger et al., 16033 (OSC)
BB	Breitenbush	44°48'35.9"N 121°52'38.0"W	Dillenberger 16018-2 (OSC)
BC	Belknap Crater	44°10'58.0"N 121°54'24.5"W	Dillenberger et al., 16021 (OSC)
BF	Butte Falls	42°29'23.0"N 122°25'16.3"W	Dillenberger et al., 16034 (OSC)
BM	Big Marsh Creek	43°27'56.3"N 121°56'44"W	Abair 16081500SB (OSC)
CL	Crater Lake	42°53'0.0"N 122°12'53.0"W	Dillenberger et al., 16031 (CRLA)
CR	Canyon Road	43°54'17.3"N 122°5'3.3"W	Dillenberger 16030 (OSC)
CW	Clearwater	43°14'48.1"N 122°17'42.7"W	Dillenberger 16019 (OSC)
DL	Diamond Lake	43°10'58.2"N 122°8'25.3"W	Dillenberger 16020 (OSC)
EL	Elk Lake	43°58'42.9"N 121°48'32.2"W	Dillenberger & Halse 16017 (OSC)
IT	Iturup Island	44°47'48.0"N 147°11'34.2"E	PI 641091*
LC	Lost Creek	44°34'39.6"N 122°7'20.5"W	Dillenberger et al., 16010 (OSC)
LL	Lost Lake	44°25'52.3"N 121°54'45.1"W	Dillenberger et al., 16011 (OSC)
LP	Lost Prairie	44°24'12.1"N 122°4'35.0"W	Dillenberger et al., 16012 (OSC)
LR	Lolo Pass Road	45°25'37.9"N 121°47'48.9"W	Dillenberger et al., 16022 (OSC)
MH	Mount Hood	45°19'18.0"N 121°42'15.7"W	Dillenberger et al., 16024 (OSC)
OL	Odell Lake	43°35'8.0"N 122°0'33.9"W	Dillenberger & Halse 16016 (OSC)
SH	Mount St. Helens	46°18'11.2"N 122°20'16.2"W	Dillenberger et al., 17041 (OSC)
TR	Tye River	47°45'0.0"N 121°6'18.0"W	Zika 27878 (OSC)
UC	Union Creek	42°54'21.6"N 122°18'23.7"W	Dillenberger et al., 16032 (OSC)
WREN2.4	Wren	44°35'16.8"N 123°25'37.2"W	Bridge Cook (OSC 37945)

Note: *Living collection at the National Clonal Germplasm Repository, Corvallis, OR.

5.1 and 200 ng ds-DNA, with 30 cycles of 30 s sonication and 30 s pause. Final libraries for sequencing were a 50/50 mix of target capture Hyb-Seq libraries and whole genome libraries. Target capture data are not discussed in this paper, and methods and results are only described for the whole genome libraries. Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs Inc., Ipswich, Massachusetts, USA). We used 50 µL of sheared DNA for library preparation. First, we end repaired and ligated with universal adaptors. Adaptor-ligated DNA was cleaned up with AMPure XP Beads (Beckman Coulter, Brea, California, USA), without size selection. The PCR was carried out using dual index primers (NEBNext Multiplex Oligos for Illumina) providing 96 unique index combinations for 8 to 12 cycles. PCR products were purified with AMPure XP Beads. The final libraries with individual indices were quantified using Qubit

2.0 and 4200 TapeStation (Agilent Technologies, Santa Clara, California, USA).

All 96 samples were run together on a single lane of an Illumina HiSeq 3000 at the Center for Genome Research and Biocomputing at Oregon State University.

Data analysis

To obtain plastid genome sequences, raw reads of all samples were mapped to a reference plastome sequence of *Fragaria virginiana* subsp. *virginiana* (sample O477; GenBank accession number NC_019602; Salamone et al., 2013) with only one copy of the inverted repeat region. For mapping, we used bbmap.sh v.36.92 of the BBTools software package (Bushnell, 2014). Sequences were trimmed based on quality on both sides (default phred quality score 6 used), minimum identity to the reference was set to 95%. The number of reads was reduced to 100,000 with reformat.sh v.36.92 (BBTools software package) for each sample to generate ~100× coverage and were saved in bam-format. Consensus sequences of mapped reads were aligned in Geneious v.9.0.5 (Kearse et al., 2012).

The quality of the resulting sequences was checked by annotating the sequences based on an annotated genome downloaded from GenBank (*F. chiloensis* HM1, NC_019601; Salamone et al., 2013). Coding regions were translated into amino acids and genes were checked for additional stop codons disrupting the gene. Although gene loss by pseudogenization might occur, it is unlikely that a substantial number of genes are lost solely in one individual. Individuals with more than one gene disrupted by additional stop codons were considered to have a sequence of low quality. This threshold was chosen based on the occasional occurrence of additional stop codons in several samples, but only two samples (OL6 and UC2, sample abbreviations according to Appendix S1) had a substantially higher number of disrupted genes. Based on this criterion, these two samples were excluded from the data set.

Phylogenetic relationships using the complete data set were inferred using RAxML v.8.0.26 (Stamatakis et al., 2008) with the fast bootstrapping algorithm, the GTR+ Γ substitution model, and automatic halt of bootstrapping based on the auto-MRE criterion with a maximum number of 1000 replicates. *Dasiphora floribunda* was used as the outgroup, based on its close relationship to *Fragaria* (Lundberg et al., 2009). This analysis was complemented by a maximum parsimony analysis performed in PAUP* v.4.0a152 (Swofford, 2002). We used 1000 random addition replicates and tree bisection and reconnection (TBR) branch swapping. Bootstrap support was calculated with PAUP* using 1000 replicates and the same settings as in the heuristic search. To reduce time to conduct analyses, we constructed a smaller data set representing one sample of the octoploid populations and three samples of *F. cascadiensis* clades found in the phylogenies of the complete data set. This reduced data set was analyzed with RAxML and PAUP*, as described above, and it was used for the following dating analysis.

The phylogeny of the reduced data set was dated using BEAST v.1.8.4 (Drummond et al., 2012). The substitution model used was GTR+ Γ +I. A relaxed, uncorrelated, lognormal clock was used, together with a birth-death process prior to inferring branch lengths. Analysis was started with a random tree. The crown node of *Fragaria* was calibrated to a minimum age of 2.9 million years (based on a fossil achene, see Liston et al., 2014) using a mean of 0.01 and a standard deviation of 1.0 in a lognormal distribution. To ensure that the effective sample size was >200, the analysis was

carried out over 100 million generations, sampling trees every 1000 generations, and discarding the first 10,000 trees as burn-in. Trace file was checked with Tracer v.1.5 (Rambaut and Drummond, 2007) and trees were annotated using TreeAnnotator v.1.5.4 (Drummond and Rambaut, 2007).

Inheritance of plastids

To test whether plastids are strictly maternally inherited in *F. cascadiensis*, or if paternal leakage occurs, we used the data set of Wei et al. (2017). We compared the plastid sequences of the parents of the experimental cross used in Wei et al. (2017), and found two unambiguous single nucleotide polymorphisms (SNPs) discriminating the two parental haplotypes. The first SNP was found in the intergenic spacer *psaA-ycf3* at alignment position 43,511 (full data set) and is a C/T transition. The second SNP was found in the *rps15-ycf1* intergenic spacer at alignment position 126,434 (full data set) and is a A/C transversion. We searched for both variants of both SNPs in the cleaned and merged sequence data (described in Wei et al., 2017; raw sequence data available at the NCBI sequence read archive: PRJNA382418) of the 45 offspring individuals of the experimental cross. We searched for the SNP and 10 bp to each side of the SNP, as well as for the complement sequences of these 21 bp. If plastid inheritance is strictly maternal, all offspring individuals should exclusively have the maternal SNP variants.

Shimodaira-Hasegawa-test

The Shimodaira-Hasegawa-test (SH-test; Shimodaira and Hasegawa, 1999) was used to test if the best tree found with RAxML for the reduced data set is significantly more likely than alternative trees containing constrained clades. First, RAxML was used to estimate trees containing constrained clades with the same settings as described for the reduced data set (tree files available at TreeBASE: <http://purl.org/phylo/treebase/phyloids/study/TB2:S22240>). We tested three different types of constrained clades using the SH-test implemented in RAxML with parameter optimization for each test. First, the four cascadiensis-clades (1, 2, 4, and 6) were individually constrained to form a clade with *F. vesca* subsp. *bracteata* individuals from Oregon (MRD population, Appendix S1; four independent SH-tests). These individuals represent the Pacific Coast clade of *F. vesca* subsp. *bracteata* that is resolved as sister to the octoploids in previous plastid-based phylogenetic analyses of *Fragaria* (Njuguna et al., 2013; Govindarajulu et al., 2015). Second, in a similar way, the four cascadiensis-clades were individually constrained to form a clade with *F. iinumae* (four independent SH-tests). Third, all four cascadiensis-clades were forced to form a single cascadiensis-clade (one SH-test). Only samples of *F. cascadiensis* and associated nonaploids were included in the constrained clades. In total, nine SH-tests were performed.

RESULTS

Sequencing statistics and coverage

Sequencing of 96 samples on a single HiSeq 3000 Illumina lane yielded a total of 299,064,448 reads, with 3,115,255 reads on average per sample. Read numbers for individual samples ranged from a minimum of 410,677 reads (sample LP6) to a maximum of 15,186,842 reads (sample AC).

Coverage of the plastome varied with total read number. For sample LP6 with the lowest number of total reads, 139,430 reads were mapped to the plastome. For sample AC with the highest total read number, 2,519,252 reads were mapped. Coverage was reduced to ~100× by using only 100,000 reads of each sample.

Flow cytometry

Of the 90 specimens believed to be *F. cascadiensis* based on morphology (Appendix S1), 81 were confirmed to be decaploid (10×), one was octoploid (8×: BB5) and eight were nonaploid (9×: BB1, BF1–5, EL2–3). Mixed ploidy populations were Breitenbush (BB: 8×, 9×, and 10×), Elk Lake and Butte Falls (EL, BF: 9× and 10×). The samples of the other 12 populations were all decaploid (Fig. 1). The six samples believed to be octoploid at collection were indeed confirmed to be 8×.

Phylogenetic relationships

The plastid phylogeny based on 537 parsimony informative sites showed that most polyploid taxa (all except *F. chiloensis*) were not monophyletic (Fig. 2, Appendix S2; tree files available at TreeBASE: <http://purl.org/phylo/treebase/phyloids/study/TB2:S22240>). Detailed support information for all phylogenies can be found in Appendix S2–S4. *Fragaria virginiana* subsp. *glauca* has three distinct lineages, with three samples in two different clades of subsp. *virginiana* (clades 3 and 5) and one with *F. cascadiensis* (clade 2; Fig. 2). *Fragaria virginiana* subsp. *virginiana* can be found in two clades (3 and 5) and forms a paraphyletic group. *Fragaria virginiana* subsp. *platypetala* is represented in the clades 6, 7, and 8 (Fig. 2, Appendix S2). It has affinities to *F. chiloensis* (clade 8), *F. cascadiensis* (clade 6) and *F. xananassa* subsp. *cuneifolia* (clades 6 and 7). *Fragaria xananassa* subsp. *cuneifolia* can be found in three clades (1, 6, and 7), with the population WREN that is present in the distantly related

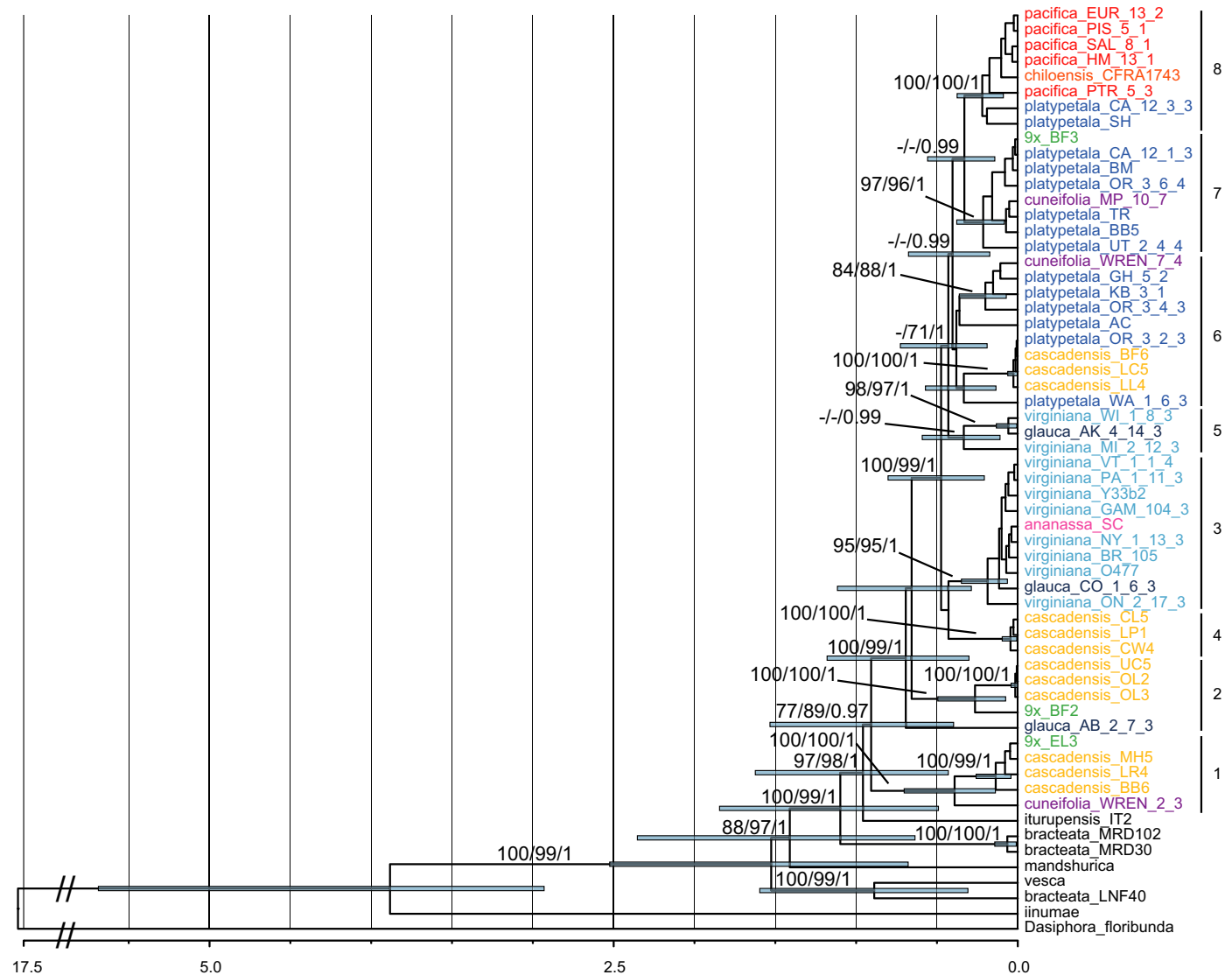


FIGURE 2. Maximum clade credibility tree of the reduced *Fragaria* plastid data set. Values above branches are maximum parsimony/maximum likelihood bootstrap values/posterior probability values, only bootstrap values > 70 and posterior probabilities > 0.95 are shown. Node bars indicate 95% confidence interval of age estimates in million years. Scale at basal branch not proportional (//). Age estimates and support values within clades (numbered according to Appendix S2) not shown. Colors indicate different taxa.

clades 1 and 6 (Fig. 2). The samples are closely related to *F. virginiana* subsp. *platypetala* (clades 6 and 7), and *F. cascadenensis* (clade 1). *Fragaria cascadenensis* is present in four clades in the plastid phylogeny. Individuals of *F. cascadenensis* can be found in the clades 1, 2, 4, and 6. Clade 1 contains samples of *F. ×ananassa* subsp. *cuneifolia*; clade 4 harbors three nonaploid samples and one sample of *F. virginiana* subsp. *glauca*; clade 4 is in different unsupported/weakly supported positions in the different phylogenies; and the samples of *F. cascadenensis* in clade 7 are related to *F. virginiana* subsp. *platypetala* (Fig. 2).

Phylogenetic dating

The results of the phylogenetic dating analysis are shown in Fig. 2 and Appendix S4. With the origin of *Fragaria* dated to a minimum of 2.9 million years ago (mya), the crown age (mean age [95% confidence interval]) of the polyploids is 0.96 mya [0.43–1.62 mya]. The crown ages of the four *cascadenensis* (sub-)clades were: 0.14 mya [0.04–0.26 mya] for clade 1, 0.02 mya [0–0.04 mya] for clade 2, 0.04 mya [0.01–0.1 mya] for clade 4, and 0.03 mya [0–0.06 mya] for clade 6 (Appendix S4).

Phylogeography of *F. cascadenensis*

The four clades for *F. cascadenensis* are not evenly distributed among the populations in the Oregon Cascades (Fig. 1). Of the 50 individuals of *F. cascadenensis* in the nine populations from Canyon Road (CR; Fig. 1) north to Lolo Pass Road (LR), 46 belong to clade 1, three to clade 6, and one to clade 4. In the six southern populations from Odell Lake (OL; Fig. 1) south to Butte Falls (BF), nine individuals of *F. cascadenensis* belong to clade 1, four to clade 2, 13 to clade 4, and three to clade 6. In the Union Creek (UC) population, all four clades can be found, whereas in the northern half of the Cascades a maximum of two clades are present in a single population. Genetic diversity is considerably higher in the southern part of the Oregon Cascades.

Inheritance of plastids

Coverage for the two regions containing the SNPs in the offspring of the experimental cross of Wei et al. (2017) was very different. The target region in the *psaA-ycf3* spacer had an average coverage of only 1.3× (and four individuals with no coverage) and was therefore excluded from further analysis. The target region in the *rps15-ycf1* spacer had an average coverage of 32× with a minimum coverage of 10×. Of the 45 offspring individuals, 34 individuals exclusively had the maternal SNP variant, and 11 individuals predominantly had the maternal SNP variant (95–99% of all reads; Table 3). Both parents had exclusively their own SNP variant. According to these results, low levels of paternal plastid transmission can be detected in ~25% of sampled offspring.

Shimodaira-Hasegawa-test

Of the nine SH-tests, all trees with constrained groups were significantly worse ($p < 0.01$) than the best tree of the reduced set (Table 4). The resulting tree in Fig. 2 is significantly more likely than any grouping of a single *cascadenensis*-clade with *F. vesca* subsp. *bracteata* from MRD, or *F. iinumae*, or together in one clade containing all samples of *F. cascadenensis*.

DISCUSSION

Our analyses resulted in the most comprehensive phylogeny of the North American polyploid *Fragaria* including *F. cascadenensis* and a broad sampling of populations and taxa. The phylogeny is well resolved with only few nodes in the backbone that are not well supported (Fig. 2, Appendix S2). Except for *F. chiloensis*, no taxon of the North American polyploid clade is monophyletic. *Fragaria chiloensis* seems to have evolved from *F. virginiana* subsp. *platypetala*-like ancestors. Furthermore, the phylogeny shows a strong geographic pattern within *F. chiloensis*. The populations HM and SAL are the northern-most populations sampled in this study and are located at the central Oregon coast. Sister to these populations are the southern Oregon–north California populations PIS and EUR. Sister to these four populations is the sample from *F. chiloensis* subsp. *chiloensis* from Chile (map showing locations in Appendix S5). Together with the fact that the samples from the central California coast population PTR are sister to all other *F. chiloensis* samples, the origin of the South American subspecies may be localized to the northern half of the Californian coast.

In contrast, *F. virginiana* subsp. *virginiana* is found in two clades, subsp. *glauca* is found in three clades, subsp. *platypetala* in at least three clades, and *F. cascadenensis* in four independent clades. *Fragaria virginiana* subsp. *virginiana* is part of clades 3 and 5. The subspecies has at least two different plastid donors. The populations in clade 5 (from Michigan and Wisconsin) are the northern and western most populations included. The presence of a sample from New York, and the Alaskan subsp. *glauca* suggest this clade might be related to the colonization of previously glaciated areas (Dyke et al., 2003). However, other individuals from glaciated northeastern North America are in the separate clade 3.

Fragaria virginiana subsp. *glauca* is polyphyletic in this phylogeny, the three included samples are located in three different clades (2, 3, and 5; Fig. 2). The sampling of subsp. *glauca* needs to be increased to cover the entire range (Staudt, 1999) and additional clades might be found. In contrast to subsp. *glauca*, subsp. *platypetala* is paraphyletic and could have had a single origin. *Fragaria ×ananassa* subsp. *cuneifolia* is a hybrid of subsp. *platypetala* and *F. chiloensis* (Govindarajulu et al., 2013; Salamone et al., 2013). In this case, it is plausible that samples of subsp. *cuneifolia* are found nested within subsp. *platypetala*. The only problematic case is clade 6 containing samples of *F. cascadenensis*. Explanations are discussed below for *F. cascadenensis*.

There are two different explanations that could account for the patterns found in the octoploid taxa. The first explanation is incomplete lineage sorting (Maddison and Knowles, 2006). The subspecies could have evolved from a pool of haplotypes that is still detectable in these taxa. The second explanation is that the subspecies of *F. virginiana* were originally monophyletic or paraphyletic, but hybridized after subspecies formation and exchanged genetic material including plastids. All subspecies of *F. virginiana* overlap in their distribution with at least one other subspecies (Staudt, 1999), enabling hybridization. Neither explanation excludes the other, and the pattern found could be the result of a mixture of these events. Although paternal transmission of plastids could reduce the number of origins needed to explain the genetic pattern, multiple origins of most octoploid taxa are still required. If only two clades of a taxon are found in the phylogeny, these could represent the maternal and paternal plastid haplotypes that were transmitted in the origin of the taxon. However, for all octoploid taxa with more than

TABLE 3. Number of parent-specific plastid sequence reads in the offspring of the experimental cross of Wei et al. (2017).

Individual*	Total no. of reads of the maternal plastid variant	Total no. of reads of the paternal plastid variant	% of specifically maternal plastid reads	% of specifically paternal plastid reads
CASXA2015.1	129	0	100	0
CASXA2015.2	99	2	98	2
CASXA2015.4	88	2	98	2
CASXA2015.5	72	0	100	0
CASXA2015.6	101	1	99	1
CASXA2015.7	64	0	100	0
CASXA2015.10	66	0	100	0
CASXA2015.12	172	1	99	1
CASXA2015.13	47	0	100	0
CASXA2015.14	26	0	100	0
CASXA2015.15	80	0	100	0
CASXA2015.16	63	0	100	0
CASXA2015.19	61	0	100	0
CASXA2015.20	56	2	97	3
CASXA2015.21	39	0	100	0
CASXA2015.22	84	0	100	0
CASXA2015.23	148	0	100	0
CASXA2015.25	94	0	100	0
CASXA2015.26	74	0	100	0
CASXA2015.27	56	1	98	2
CASXA2015.28	52	1	98	2
CASXA2015.29	49	0	100	0
CASXA2015.30	59	0	100	0
CASXA2015.31	14	0	100	0
CASXA2015.33	19	0	100	0
CASXA2015.34	22	0	100	0
CASXA2015.35	26	0	100	0
CASXA2015.36	31	1	97	3
CASXA2015.37	88	0	100	0
CASXA2015.38	56	0	100	0
CASXA2015.40	48	0	100	0
CASXA2015.41	96	1	99	1
CASXA2015.42	29	0	100	0
CASXA2015.44	131	0	100	0
CASXA2015.46	46	1	98	2
CASXA2015.47	123	1	99	1
CASXA2015.48	89	0	100	0
CASXA2015.62	12	0	100	0
CASXA2015.64	15	0	100	0
CASXA2015.75	58	0	100	0
CASXA2015.76	61	0	100	0
CASXA2015.77	38	2	95	5
CASXA2015.81	78	0	100	0
CASXA2015.82	63	0	100	0
CASXA2015.83	78	0	100	0
Maternal parent	28	0	100	0
Paternal parent	0	10	0	100

Note: *Names identical to individuals in the sequence read archive (PRJNA382418).

two clades (*Fragaria* × *ananassa* subsp. *cuneifolia*, *F. virginiana* subsp. *platypetala*, and *F. virginiana* subsp. *glauca*) the previously discussed explanations apply.

Fragaria virginiana subsp. *virginiana* occurs from the east coast of North America extending westwards where the range of subsp. *glauca* begins (Staudt, 1999). Subspecies *glauca* occurs mainly along the Rocky Mountains and overlaps with subsp. *platypetala* in the west. *Fragaria chiloensis* in North America is restricted to the west coast (Staudt, 1999; see sampling in Appendix S5). This east-to-west directionality can also be found to some extent in the phylogeny (Appendix S2). Samples or taxa that do not fit this geographic pattern in the phylogeny (e.g., subsp.

glauca) might be explained by hybridization with other octoploids, incomplete lineage sorting, or paternal leakage of plastids.

Evolution of *Fragaria cascadenis*

As the only decaploid member of the North American polyploids, *F. cascadenis* was initially assumed to be derived from a cross of sympatric *F. virginiana* subsp. *platypetala* and *F. vesca* subsp. *bracteata* (Hummer, 2012). In contrast, Wei et al. (2017) did not support this hypothesis, and instead hypothesized an extinct relative of *F. iinumae* and an ancestral octoploid were the progenitors of

TABLE 4. Results of the Shimodaira-Hasegawa-tests.

Constrained Group	Likelihood of the best tree (Fig. 2)	Likelihood of the constrained tree	Standard deviation	p-value ²
clade 1+MRD ¹	−198368.391675	−236.059716	41.621224	< 0.01
clade 2+MRD	−198232.529082	−100.197123	26.867659	< 0.01
clade 4+MRD	−198299.591877	−167.259917	34.327745	< 0.01
clade 6+MRD	−198480.163399	−347.831440	48.334129	< 0.01
clade 1+F. <i>iinumae</i>	−198486.511622	−354.179662	42.708136	< 0.01
clade 2+F. <i>iinumae</i>	−198369.000765	−236.668805	35.014235	< 0.01
clade 4+F. <i>iinumae</i>	−198411.010872	−278.678912	38.221835	< 0.01
clade 6+F. <i>iinumae</i>	−198520.510244	−388.178285	45.052085	< 0.01
clade 1+2+4+6	−198722.494837	−590.162877	66.243903	< 0.01

Notes: ¹MRD: *F. vesca* subsp. *bracteata* population MRD. ²Constrained trees significantly worse if $p < 0.01$.

F. cascadiensis. The position of the four clades of *F. cascadiensis* within the North American polyploids, but largely distantly related from *F. virginiana* subsp. *platypetala*, is consistent with the conclusion of Wei et al. (2017), but also showed evidence of at least four plastid donors (Fig. 2, Appendix S2). There is evidence that hybridization of taxa with different ploidy levels is more likely to occur when the paternal parent has the higher ploidy level (Kreiner et al., 2017). *Fragaria cascadiensis* is the result of a cross (or crosses) between an octoploid parent and a diploid *F. iinumae*-like genome (Wei et al., 2017). The octoploid parent(s) of *F. cascadiensis* presumably had a *F. vesca* subsp. *bracteata*-like plastome, because all extant octoploids share plastids that are derived from *F. vesca* subsp. *bracteata* (Fig. 2; Govindarajulu et al., 2015). None of the *F. cascadiensis* plastome clades show evidence (SH-test, $p < 0.01$) of recent common ancestry with *F. iinumae*, the closest relative of its diploid ancestor, nor with *F. vesca* subsp. *bracteata*—including three individuals from Oregon (Fig. 2, Appendix S2).

We documented paternal transmission of plastids in *F. cascadiensis*. Thus, we cannot exclude the possibility that the octoploid ancestor of *F. cascadiensis* acted as paternal parent in the cross with the diploid *F. iinumae*-like ancestor. However, considering the limited degree of paternal inheritance (Table 3), maternal inheritance of plastids is more likely according to our results. In conclusion, the plastid genome was inherited from the octoploid ancestor, which was more likely the maternal parent. This contradicts previous evidence that hybridization is more likely with the paternal parent having the higher ploidy level (Kreiner et al., 2017). If the *F. iinumae*-like ancestor of *F. cascadiensis* ever transmitted plastids either as maternal or paternal parent, there is no evidence left in the extant populations of *F. cascadiensis*.

There are two different explanations for the existence of four independent *cascadiensis*-clades in the phylogeny. The first explanation is the recurrent origin of *F. cascadiensis*. This could have taken place in independent populations, or perhaps a single population containing all four haplotypes found in *F. cascadiensis*. Based on the position of clade 1 (sister to almost all North American octoploid haplotypes; Fig. 2, Appendix S2), the haplotype diversity in *F. cascadiensis* is probably not a result of recent hybridization events, but can be dated back to the time of the diversification of the octoploids. The exact location of this one, or multiple, populations is difficult to determine, but it could be the same location as hypothesized for the origin of the octoploids, i.e., Beringia (Liston et al., 2014). The east-west pattern of octoploid samples found for subsp. *virginiana*, subsp. *platypetala* and *F. chiloensis* is likely a coincidence that was caused by lineage sorting after the dispersal to their present range. The alternative explanation assumes that the decaploid *F. cascadiensis* evolved once and additional plastid haplotypes were introduced later. A single origin in

combination with paternal leakage can only explain two haplotypes within the offspring, but in our case four haplotypes were found. We consider the possibility of a single origin of *F. cascadiensis* to be unlikely. This would imply that *F. cascadiensis* crossed repeatedly with octoploid plants with different plastid genomes to introduce their haplotypes into *F. cascadiensis*. To arrive at decaploid *F. cascadiensis* plants with captured plastids it would require that nonaploid hybrids backcrossed to decaploid *F. cascadiensis*. However, this scenario harbors the problem that meiosis is most likely disturbed due to the unbalanced number of chromosomes in these hybrids (Norrman et al., 1997). Still, the possibility exists that *F. cascadiensis* gametes by chance had the appropriate chromosome number. This scenario would also require that *F. cascadiensis* was in contact with all relevant haplotypes. A likely place of contact is Beringia where the hypothesized origin of the polyploids can be found (Liston et al., 2014). Fertility of nonaploid *Fragaria* hybrids is unknown. Thus, we cannot rule out the possibility of a rare event of viable pollen/ovule formation and transmittance of plastids into *F. cascadiensis* from octoploid relatives.

It is also possible that a combination of these two scenarios took place. While some or most haplotypes became part of the haplotype pool of *F. cascadiensis* during its origin by multiple hybrid formation with different plastid donors, the *platypetala*-like haplotype of the clade 6 (Fig. 2) could be the result of chloroplast capture that happened later in the evolution of *F. cascadiensis*. *Fragaria cascadiensis* and *F. virginiana* subsp. *platypetala* hybridize where they co-occur. We found eight plants that turned out to be nonaploid. One plausible explanation for their origin is that these plants are hybrids of the sympatric taxa *F. cascadiensis* and *F. virginiana* subsp. *platypetala*. The other possible explanation is that these plants are hybrids of octoploid *F. virginiana* subsp. *platypetala* providing unreduced (8×) gametes and *F. vesca* subsp. *bracteata* providing normally reduced (1×) gametes. The first explanation is more likely, because *F. cascadiensis* and *F. virginiana* subsp. *platypetala* co-occurred in all three populations with nonaploid plants, but *F. vesca* subsp. *bracteata* was found in only one of these populations. The above scenarios may be differentiated by analyzing the nuclear genome, where traces of introgression from octoploids should be present, if hybridization played a role in recent evolution of *F. cascadiensis*.

Wei et al. (2017) used linkage mapping to determine the ancestors of *F. cascadiensis*. Both parents used in their cross are part of clade 1 (*cascadiensis_OR_5_3_2* and *cascadiensis_OR_5_9_2*; Appendix S2). The results of Wei et al. (2017), confirmed by our study, may only apply to clade 1. The recurrent formation of *F. cascadiensis* implies the possibility that different lineages of *F. cascadiensis* could have different genome compositions. This needs to be tested for the remaining three lineages found in the plastid phylogeny (clades 2, 4, and 6; Fig. 2).

Multiple origins of polyploids have been inferred for most polyploid plant species that have been examined (Soltis and Soltis, 1999; Soltis et al., 2014). Classic examples of multiple origin of polyploids (e.g., in *Tragopogon* and *Draba*) show morphological differences among separate origins (Soltis and Soltis, 1999). Irrespective of the actual evolutionary scenario, the age estimate indicated that *F. cascadenensis* is not a very young hybrid that evolved in the past couple of thousands of years, but is as old as all other North American *Fragaria* polyploids (see also Wei et al., 2017).

The only other natural decaploid species in *Fragaria*, *F. iturupensis*, is endemic to Iturup Island in the southern part of the Kuril Archipelago northeast of Japan. *Fragaria iturupensis* is sister to all North American polyploids. Without increased sampling, it remains unknown if *F. iturupensis* shows a similar pattern of multiple origins as other polyploids. These two decaploids could have evolved in Beringia and then moved southwards on different sides of the Pacific Ocean and became isolated (Wei et al., 2017). The High Cascades and the Kuril Archipelago are both of recent volcanic origin, suggestive of similar habitat requirements for the two species.

An alternative scenario of the evolution of the polyploid clade (North American polyploids + *F. iturupensis*; Appendix S2) can be inferred from the distribution of ploidy on the phylogeny. Most of the samples in the basal clades of the polyploids are decaploid (*F. iturupensis*, clade 1, clade 2; Fig. 2, Appendix S2). This could imply that the ancestor of all polyploids was decaploid, or (more likely) the octoploid that gave rise to the decaploids also contributed to the extant octoploid taxa. If decaploidy was ancestral, the octoploids would then represent the result of a genome reduction. Balanced reduction of the genome by one subgenome, keeping integrity of the remaining four subgenomes appears to be unlikely. Therefore, we assume that decaploidy was derived from octoploid ancestors. Whether it occurred once in a common ancestor of *F. cascadenensis* and *F. iturupensis*, or independently in these species, remains to be determined.

Phylogeography and age of *Fragaria cascadenensis*

Fragaria cascadenensis is currently restricted to the Oregon Cascades. Morphologically similar plants from Washington, California and other mountain ranges in Oregon were shown to be octoploid (Hummer, 2015). A plausible region for the origin of *F. cascadenensis* is Beringia, where the octoploids most likely evolved (Liston et al., 2014). In this region, all necessary progenitors of *F. cascadenensis* were presumably present, including the donor of the additional *F. iinumae*-like B subgenome. In this case, it is difficult to explain why *F. cascadenensis* is restricted to the Oregon Cascades and has not been found in other parts of the Cascades or the Rocky Mountains that harbor suitable habitats for strawberries. Although these regions are suitable for closely related *Fragaria* taxa (*F. virginiana* subsp. *glauca* and subsp. *platypetala*), biotic differences might be strong enough to prevent *F. cascadenensis* in these regions.

An alternative, very unlikely place of origin is the Oregon Cascades. This would explain the present distribution, but would assume that the progenitors of *F. cascadenensis* were found in this region in the past. There is no evidence that *F. iinumae* and close relatives, now restricted to eastern Asia (Liston et al., 2014), were ever found in North America.

The results of our age estimation show that the polyploids are quite young (0.43–1.62 million years; Appendix S4) compared to the rest of the genus. The origin of the octo- and decaploid taxa

is placed in the Pleistocene in coincidence with the presence of Beringia (Hopkins, 2013). Two other recent studies published age estimates of *Fragaria* that are considerably older than ours: Qiao et al. (2016) estimated that *Fragaria* originated 7.99 mya, but their calibration was based on an age estimate of the split of Rosoideae and Maloideae+Prunoideae, and did not include the fossil used here. Xiang et al. (2016) estimated the age of *Fragaria* to ~12 million years old, using the program r8s (Sanderson, 2003). Differences in age estimates can be explained by methodological differences in these analyses incorporating different calibrations and programs. Furthermore, Xiang et al. (2016) placed the *Fragaria* fossil in their tree at the split of *Fragaria* and *Comarum* L. (a stem calibration) and this is dated to ~40 mya, although the fossil is dated to 2.9 mya. This suggests a potentially inappropriate calibration of this node in their study. The possibility of an older age of *Fragaria* would provide more time for migration and contact of species, and opens the possibility of additional scenarios for the origin of *F. cascadenensis*.

Although *F. cascadenensis* has four major plastid haplotypes, these are not occurring in the same frequency, nor are they evenly distributed among populations. A total of 57 individuals of *F. cascadenensis* belong to clade 1, while only 24 individuals belong to the clades 2, 4 and 6 taken together. Furthermore, clade 1 is the dominant clade in the northern half of the Oregon Cascades and the other clades are mainly found in the southern half (Fig. 1). In one population (UC), all four clades are found among the six individuals. Regarding the decaploid samples, clade 2 is restricted to the southern half of the Oregon Cascades. The *F. cascadenensis* samples of the largest cascadenensis-subclade (part of clade 1) diversified first (0.04–0.26 mya; Fig. 2, Appendix S4), while the middle-sized clade 4 diversified between 0.01–0.10 mya and the smallest cascadenensis-subclades (part of clades 2 and 6) diversified between 0–0.04 and 0–0.06 mya.

Fragaria cascadenensis is the only known vascular plant species that is widespread in the Oregon Cascades, but restricted to this ecoregion (Oregon Flora Project, 2004; Meyers et al., 2015). The approximately 12 other Cascades endemics have a restricted distribution in a small part of the range. Dyke et al. (2003) showed that the Cascades in Oregon were most likely glaciated during the last glacial maximum (LGM). Although climatic conditions might have been more favorable during the LGM for a species of higher altitudes, it is clear that *F. cascadenensis* cannot have persisted in the range where it is found today. A number of refugia have been identified in previous studies of the Pacific Northwest region (e.g., Roberts and Hamann, 2015) and it is likely that *F. cascadenensis* survived the LGM in one or more of these. For tree species, two refugia adjacent to the Oregon Cascades have been identified (Roberts and Hamann, 2015) that could also have harbored *F. cascadenensis*. The first refugium is located in the Willamette Valley, west of the Oregon Cascades, and relatively small and species-poor compared to other refugia found in that study. The second, larger refugium is the Klamath Mountains, southwest of the Oregon Cascades. *Fragaria cascadenensis* could have survived in one or both of these refugia. Considering the present distribution of *Fragaria* species in these regions, octoploids can only be found in the Willamette Valley, while diploids are found in both regions (Oregon Flora Project, 2004), indicating perhaps a higher suitability of the Willamette Valley as refugium of *F. cascadenensis*. The Cascades could have been recolonized after the LGM from both refugia—the northern part from the more northern Willamette Valley, the

southern part from the Klamath Mountains. The Willamette Valley as a smaller refugium may have harbored a smaller population of *F. cascadiensis*, increasing the effect of genetic drift and bottlenecks, explaining the lower number of haplotypes compared to the populations in the south. If *F. cascadiensis* only survived in the Klamath Mountains, recolonization of the Cascades could also have led to the present distribution of genetic diversity. While moving from south to north, the populations of *F. cascadiensis* had to face repeated founder effects and bottlenecks, which decreased genetic diversity. Survival of *F. cascadiensis* near the southern part of the Cascades is supported by the fact that refugial areas often harbor higher diversity (Hewitt, 2004). Another explanation for high diversity in the south is the possibility that different lineages met in a contact zone. *Fragaria cascadiensis* could have been distributed over a larger number of refugia during the LGM and diversified there, subsequently meeting in the southern Cascades. This scenario is less likely because we would expect a widespread *F. cascadiensis* during the LGM to have some relictual populations in other mountain ranges of North America, e.g., the Sierra Nevada or Rocky Mountains.

Notes on *Fragaria* × *ananassa*

The two subspecies of *Fragaria* × *ananassa* have different origins. Subspecies *cuneifolia* is a natural hybrid occurring in western North America, subspecies *ananassa* is an artificial hybrid that is widely cultivated (Liston et al., 2014). The subsp. *ananassa* cultivar ‘Sweet Charlie’ was included, showing a plastid ancestry of *F. virginiana* subsp. *virginiana*. *Fragaria chiloensis* is well-documented as the maternal ancestor of cultivated strawberries (Lee, 1966), suggesting that either subsequent introgression with *F. virginiana* subsp. *virginiana* has occurred in the ancestry of at least some North American cultivars introducing *F. virginiana* subsp. *virginiana*-like plastid haplotypes, or paternal leakage in the original crosses resulted in the presence of *F. virginiana* subsp. *virginiana*-like plastid haplotypes in subsp. *ananassa*. Examination of plastid phylogeny in more cultivars will be required to resolve this. The position of the samples *cuneifolia_WREN_2_3* and *cuneifolia_WREN_2_4* indicates the existence of haplotypes in *F. virginiana* subsp. *platypetala* that were not sampled, because *F. virginiana* subsp. *platypetala* is one of the parental taxa of *F. ×ananassa* subsp. *cuneifolia*. Govindarajulu et al. (2013) and Salamone et al. (2013) used SNPs in two plastid genes to identify maternal lineages in *F. ×ananassa* subsp. *cuneifolia* to identify lineages that belong to *F. virginiana* and *F. chiloensis*. By increasing the sampling in our study, it turned out that these SNPs did not discriminate between *virginiana*-like and *chiloensis*-like haplotypes, but are diagnostic for different clades of *F. virginiana* subsp. *platypetala* (clades 6 and 7). Furthermore, none of the sampled *F. ×ananassa* subsp. *cuneifolia* individuals has a plastid haplotype similar to *F. chiloensis*, although *F. chiloensis* is considered to be one of the hybrid’s parents. Either hybridization occurred unidirectionally with *F. virginiana* subsp. *platypetala* as maternal parent (and no paternal leakage), or plants with *chiloensis*-like haplotypes were not sampled or are extinct.

Future work and possible improvements

This study provides novel insight into the evolution of the North American polyploid species of *Fragaria*. Nevertheless, several points need further attention. The sampling is not complete.

Fragaria chiloensis subsp. *sandwicensis* (O. Deg. & I. Deg.) Staudt from Hawaii and *F. virginiana* subsp. *grayana* (E. Vilm. ex J. Gay) Staudt from southeastern United States are missing. The sampling for *F. chiloensis* subsp. *pacifica* north of Oregon needs to be improved, as well as the sampling of *F. virginiana* subsp. *glauca*. Also, the South American *F. chiloensis* subsp. *chiloensis* is not well represented and additional samples should be sequenced to confirm our results. For *F. virginiana*, the Canadian part of the distribution is underrepresented and could be improved.

Another point that should get attention is the evolution of *F. iturupensis*. A population sampling of this species could help to answer the question of ploidy in this species (8×: Staudt, 1973; 10×: Hummer et al., 2009) and results could be compared to *F. cascadiensis*.

For the present data set, analysis of the nuclear genome could help to distinguish between incomplete lineage sorting and hybridization in the evolution of the octoploids. Furthermore, nuclear data provides insights into both parental lineages and might reveal evidence for additional ancestral lineages not found in the present study. In addition, data from the nuclear genome will be required to clarify the ancestry of the nonaploid individuals.

CONCLUSIONS

This is the first phylogenetic study of the plastid genome in *Fragaria* including *Fragaria cascadiensis*. We showed that intensive sampling of a species and sequencing of the whole plastome can provide new insights into the evolution of a species. For *F. cascadiensis*, the presence of all four haplotypes in a single extant population suggests these origins could potentially have occurred within a single ancestral population. Genetic diversity is considerably lower in the northern half of its distribution and can be explained by range shift during the Quaternary climatic oscillations (Hewitt, 2004). Nuclear data is needed to investigate the origin and evolution of this species further and to answer the question if additional genetic lineages were involved.

ACKNOWLEDGEMENTS

The authors thank the USDA Forest Service and Crater Lake National Park for granting collection permits, and Jennifer Beck for assistance at Crater Lake National Park. We thank Kristen Finch and Rich Cronn for laboratory assistance, and the OSU Center for Genome Research and Biocomputing for Illumina sequencing and computational support. We thank Kim Hummer (USDA National Clonal Germplasm Repository) for providing leaf tissue of *F. iturupensis*, fieldwork assistance, and discussions. We thank two anonymous reviewers for helpful comments on the manuscript. MSD thanks the German Research Foundation (DFG; DI 2144/1-1) for financial support. The US National Science Foundation (DEB 1241217 to AL & DEB 1020523 to TLA) also provided financial support.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

- Bernhardt, N., J. Brassac, B. Kilian, and F. R. Blattner. 2017. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology* 17: 141.
- Brochmann, C., P. S. Soltis, and D. E. Soltis. 1992. Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. *Nordic Journal of Botany* 12: 257–272.
- Bushnell, B. 2014. BBMap short read aligner. Computer program and documentation distributed by the author, website: <https://sourceforge.net/projects/bbmap/> [accessed 02 February 2017].
- Corriveau, J. L., and A. W. Coleman. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperms. *American Journal of Botany* 75: 1443–1458.
- Davis, T. M., M. E. Shields, A. E. Reinhard, P. A. Reavey, J. Lin, H. Zhang, and L. L. Mahoney. 2010. Chloroplast DNA inheritance, ancestry, and sequencing in *Fragaria*. *Acta Horticulturae* 859: 221–228.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology* 7: 214.
- Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.
- Dyke, A. S., A. Moore, and L. Robertson. 2003. Deglaciation of North America. Geological Survey of Canada Open File 1574, Ottawa, Ontario, Canada.
- Ellis, J. R., K. E. Bentley, and D. E. McCauley. 2008. Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower *Helianthus verticillatus*. *Heredity* 100: 574–580.
- Fawcett, J. A., S. Maere, and Y. Van de Peer. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences, USA* 106: 5737–5742.
- Folk, R. A., J. R. Mandel, and J. V. Freudenstein. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of Angiosperms. *Systematic Biology* 66: 320–337.
- Govindarajulu, R., A. Liston, and T.-L. Ashman. 2013. Sex-determining chromosomes and sexual dimorphism: insights from genetic mapping of sex expression in a natural hybrid *Fragaria* × *ananassa* subsp. *cuneifolia*. *Heredity* 110: 430–438.
- Govindarajulu, R., M. Parks, J. A. Tennessen, A. Liston, and T.-L. Ashman. 2015. Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany* 102: 544–554.
- Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 359: 183–195.
- Hopkins, D. M. 2013. Aspects of the Paleogeography of Beringia during the Late Pleistocene. In D. M. Hopkins, J. V. Matthews, C. E. Schweger, and S. B. Young [eds.], *Paleoecology of Beringia*, 3–28. Academic Press, New York, New York, USA.
- Huang, D. I., C. A. Hefer, N. Kolosova, C. J. Douglas, and Q. C. B. Cronk. 2014. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytologist* 204: 693–703.
- Hummer, K. E. 2012. A new species of *Fragaria* (Rosaceae) from Oregon. *Journal of the Botanical Research Institute of Texas* 6: 9–15.
- Hummer, K. E. 2015. The discovery and naming of the Cascade Strawberry (*Fragaria cascadenensis* Hummer). *Kalmiopsis* 21: 26–31.
- Hummer, K. E., P. Nathewet, and T. Yanagi. 2009. Decaploidy in *Fragaria iturupensis* (Rosaceae). *American Journal of Botany* 96: 713–716.
- Husband, B. C., D. W. Schemske, T. L. Burton, and C. Goodwillie. 2002. Pollen competition as a unilateral reproductive barrier between sympatric diploid and tetraploid *Chamerion angustifolium*. *Proceedings of the Royal Society of London, B, Biological Sciences* 269: 2565–2571.
- Jenkins, G., and H. Rees. 1991. Strategies of bivalent formation in allopolyploid plants. *Proceedings of the Royal Society of London, B, Biological Sciences* 243: 209–214.
- Kamneva, O. K., J. Syring, A. Liston, and N. A. Rosenberg. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evolutionary Biology* 17: 180.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
- Kreiner, J. M., P. Kron, and B. C. Husband. 2017. Evolutionary dynamics of unreduced gametes. *Trends in Genetics* 33: 583–593.
- Lee, V. 1966. Duchesne and his work. In G. M. Darrow [ed.], *Strawberry: History, breeding and physiology*, 40–72. Holt, Rinehart & Winston, New York, New York, USA. Website: http://specialcollections.nal.usda.gov/speccoll/collectionsguide/darrow/Darrow_TheStrawberry.pdf [accessed 04 October 2017].
- Liston, A., R. Cronn, and T.-L. Ashman. 2014. *Fragaria*: A genus with deep historical roots and ripe for evolutionary and ecological insights. *American Journal of Botany* 101: 1686–1699.
- Lundberg, M., M. Töpel, B. Eriksen, J. A. A. Nylander, and T. Eriksson. 2009. Allopolyploidy in *Fragariinae* (Rosaceae): Comparing four DNA sequence regions, with comments on classification. *Molecular Phylogenetics and Evolution* 51: 269–280.
- Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.
- Mallet, J. 2007. Hybrid speciation. *Nature reviews* 446: 279–283.
- Matsushima, R., Y. Hu, K. Toyoda, S. Sodmergen, and W. Sakamoto. 2008. The model plant *Medicago truncatula* exhibits biparental plastid inheritance. *Plant and Cell Physiology* 49: 81–91.
- Meyers, S. C., T. Jaster, K. E. Mitchell, and L. K. Hardison. 2015. *Flora of Oregon*. Volume 1: Pteridophytes, Gymnosperms, and Monocots. Botanical Research Institute of Texas, Fort Worth, Texas, USA.
- Mogensen, H. L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany* 83: 383–404.
- Neale, D. B., and R. R. Sederoff. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics* 77: 212–216.
- Njuguna, W., A. Liston, R. Cronn, T.-L. Ashman, and N. Bassil. 2013. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* 66: 17–29.
- Norrmann, G. A., C. L. Quarin, and K. H. Keeler. 1997. Evolutionary implications of meiotic chromosome behavior, reproductive biology, and hybridization in 6x and 9x cytotypes of *Andropogon gerardii* (Poaceae). *American Journal of Botany* 84: 201–207.
- Oregon Flora Project. 2004. Oregon plant atlas. Oregon State University, Corvallis, Oregon, USA. Website <http://www.oregonflora.org/oregonplantatlas.html> [accessed 04 October 2017].
- Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.
- Paun, O., F. Forest, M. F. Fay, and M. W. Chase. 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytologist* 182: 507–518.
- Pham, K. K., A. L. Hipp, P. S. Manos, and R. C. Cronn. 2017. A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny. *Genome* 60: 720–732.
- Qiao, Q., L. Xue, Q. Wang, H. Sun, Y. Zhong, J. Huang, J. Lei, and T. Zhang. 2016. Comparative transcriptomics of strawberries (*Fragaria* spp.) provides insights into evolutionary patterns. *Frontiers in Plant Science* 7: 1839.
- Rambaut, A., and A. J. Drummond. 2007. Tracer v1.6. MCMC Trace File Analyser. Computer program and documentation distributed by the author, website: <http://tree.bio.ed.ac.uk/software/tracer/> [accessed 10 May 2016].
- Reboud, X., and C. Zeyl. 1994. Organelle inheritance in plants. *Heredity* 72: 132–140.
- Rieseberg, L. H., and J. H. Willis. 2007. Plant speciation. *Science* 317: 910–914.
- Roberts, D. R., and A. Hamann. 2015. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society of London, B, Biological Sciences* 282: 20142903.
- Salamone, I., R. Govindarajulu, S. Falk, M. Parks, A. Liston, and T.-L. Ashman. 2013. Bioclimatic, ecological, and phenotypic intermediacy and high genetic

- admixture in a natural hybrid of octoploid strawberries. *American Journal of Botany* 100: 939–950.
- Sambrook, J., and D. W. Russell. 2001. Molecular cloning. Cold Spring Harbor Press, Cold Spring Harbor, New York, New York, USA.
- Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301–302.
- Sang, T., D. Crawford, and T. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.
- Schwarzbach, A. E., and L. H. Rieseberg. 2002. Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology* 11: 1703–1715.
- Segraves, K. A., J. N. Thompson, P. S. Soltis, and D. E. Soltis. 1999. Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology* 8: 253–262.
- Sherman-Broyles, S., A. Bombarely, and J. Doyle. 2017. Characterizing the allopolyploid species among the wild relatives of soybean: Utility of reduced representation genotyping methodologies. *Journal of Systematics and Evolution* 55: 365–376.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.
- Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202: 1105–1117.
- Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. *TREE* 14: 348–352.
- Soltis, D. E., P. S. Soltis, J. C. Pires, A. Kovarik, J. A. Tate, and E. Mavrodiev. 2004. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* 82: 485–501.
- Spooner, D. M., H. Ruess, M. Iorizzo, D. Senalik, and P. Simon. 2017. Entire plastid phylogeny of the carrot genus (*Daucus*, Apiaceae): Concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. *American Journal of Botany* 104: 296–312.
- Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology* 75: 758–771.
- Staudt, G. 1973. *Fragaria iturupensis*: A new species of strawberry from East Asia. *Willdenowia* 7: 101–104.
- Staudt, G. 1999. Systematics and geographic distribution of the American strawberry species: Taxonomic studies in the genus *Fragaria* (Rosaceae: Potentilleae). University of California Press, Berkeley, California, USA.
- Staudt, G., and K. Olbricht. 2008. Notes on Asiatic *Fragaria* species V: *F. nipponica* and *F. iturupensis*. *Botanische Jahrbücher* 127: 317–341.
- Straub, S. C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitemier, R. C. Cronn, and A. Liston. 2011. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12: 211.
- Straub, S. C. K., M. J. Moore, P. S. Soltis, D. E. Soltis, A. Liston, and T. Livshultz. 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution* 80: 165–189.
- Svab, Z., and P. Maliga. 2007. Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *PNAS* 104: 7003–7008.
- Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer, Sunderland, Massachusetts, USA.
- Szmidt, A. E., T. Aldén, and J.-E. Hällgren. 1987. Paternal inheritance of chloroplast DNA in *Larix*. *Plant Molecular Biology* 9: 59–64.
- Tennessen, J. A., R. Govindarajulu, T.-L. Ashman, and A. Liston. 2014. Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biology and Evolution* 6: 3295–3313.
- Tennessen, J. A., N. Wei, S. C. K. Straub, R. Govindarajulu, A. Liston, and T.-L. Ashman. 2017. Repeated translocation of a gene cassette drives sex chromosome turnover in strawberries. *bioRxiv* preprint, <https://doi.org/10.1101/163808>
- Wei, N., J. A. Tennessen, A. Liston, and T.-L. Ashman. 2017. Present-day sympatry belies the evolutionary origin of a high-order polyploid. *New Phytologist* 216: 279–290.
- Wendel, J. F. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* 102: 1753–1756.
- Xiang, Y., C.-H. Huang, Y. Hu, J. Wen, S. Li, T. Yi, H. Chen, et al. 2016. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution* 34: 262–281.
- Zhang, Q., Y. Liu, and Sodmergen. 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant and Cell Physiology* 44: 941–951.