Geographic patterns of genetic variation in three genomes of North American diploid strawberries with special reference to *Fragaria vesca* subsp. *bracteata*

Lauren Stanley[1], Nicole J. Forrester[1], Rajanikanth Govindarajulu[1], Aaron Liston[2], and Tia-Lynn Ashman[1,3]

[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, USA

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, 97331, USA

[3]Author for correspondence (email: tia1@pitt.edu) fax: 412-624-4759; phone: 412-624-0984

Author emails:

Lauren Stanley (lauren.stanley@uconn.edu)

Nicole J. Forrester (njf26@pitt.edu)

Rajanikanth Govindarajulu (raj2010@pitt.edu)

Aaron Liston (listona@science.oregonstate.edu)

Abstract: Geographic patterns of genetic variation in wild species reflect the interplay of ecological and evolutionary processes. We assessed genetic variation in three genomes across four North American diploid strawberry taxa, with special emphasis on the gynodioecious *F. vesca* subsp. *bracteata*. Specifically, we sequenced one chloroplast (*rpoC2*) and two mitochondrial (*atp8* and *atp8-orf225*) genes along with several nuclear microsatellite markers. In addition we assessed indicators of breeding system (pollen viability and female frequency) for all taxa. The geographic perspective on the distribution of cytoplasmic and nuclear variation revealed the genetic affiliation of the restricted taxa (*F. v.* subsp. *californica* and *F. mexicana*) with the widespread *F. v.* subsp. *bracteata* and identified a hotspot of hybridization within gynodioecious *F. v.* subsp. *bracteata*. Higher pollen viability of hermaphrodites was found in the three hermaphroditic taxa relative to the gynodioecious one. Although theoretically predicted to be associated, proportion females within *F. v.* subsp. *bracteata* populations, was not correlated with population-level genetic variation, suggesting that the history of hybridization or population size variation is more influential on the distribution of genetic variation than sex ratio in this gynodioecious species. The documented patterns of genetic variation in this complex serve as an important point of reference for future ecological and evolutionary research in diploid *Fragaria*.

Key words: chloroplast, hybridization, genetic structure, gynodioecy, mitochondria, pollen viability

Introduction

Patterns of phenotypic and genetic variation of wild species reflect their ecological and evolutionary history. Important questions that have emerged are: How is variation partitioned among and within populations for widespread species, and how differentiated and diverse are rare species relative to widespread ones (Eckert et al. 2008; Gitzendanner and Soltis 2000; Hamrick and Godt 1989)? Placing the answers to these questions in a geographic context can contribute to a landscape-level perspective on how geographic and environmental features influence gene flow, population structure and local adaptation in wild species (Manel et al. 2009).

Studying the landscape level patterns of genetic variation in all three plant genomes (nucleus, chloroplast, and mitochondrion), in particular, can provide unique insight into the distribution of genetic diversity within and among taxa. The cytoplasmic genomes (mitochondria and chloroplast) are haploid and mostly maternally inherited in flowering plants (via ovules) whereas the nuclear genome is biparentally inherited (via pollen and ovules). Thus, comparison of geographic patterns of cytoplasmic genomes to nuclear ones can reveal the spatial pattern of admixture (hybridization), as well as identify the prevailing direction of introgression via pollen or seeds among highly differentiated populations or [sub]species (Matosiuk et al. 2014; Seneviratne et al. 2012; Vekemans and Hardy 2004). Moreover, geographic patterns of nuclear genome diversity can be assessed in light of species or population attributes that are expected to affect diversity. For instance, aspects of the breeding system that enforce outcrossing, such as gynodioecy (the coexistence of females and hermaphrodites), can lead to increased nuclear genetic variation within populations and less differentiation among them than aspects of the

breeding system that promote selfing, e.g., self-compatibility and hermaphroditism (Barrett 2003). Moreover, in gynodioecious species, the frequency of females can also affect how genetic variation is partitioned within the population, i.e., population level of inbreeding. Specifically, the higher the frequency of females, the lower the overall inbreeding level (Cuevas et al. 2006; Tarayre and Thompson 1997). Thus, comparisons among populations, or among closely related taxa, that differ in breeding system can provide insight into drivers behind the patterns of genetic diversity (Hamrick and Godt 1989).

Here we report the geographic patterns of genetic variation in all three genomes of North American diploid wild strawberries (Rosaceae). There are two widespread subspecies (*Fragaria vesca* subsp. *bracteata* [Heller] Staudt, and *F. v*. subsp. *americana* [Porter] Staudt), and two more restricted taxa (*F. v*. subsp. *californica* [Chamiso and Schlecht] Staudt, and *F. mexicana* Schlecht) (Fig. 1). All are interfertile (Bors and Sullivan 2005; R. Dalton et al. unpublished data). The subspecies *F. v.* subsp. *bracteata*, in particular, is of special interest because it is the maternal genome contributor to the octoploid progenitors (i.e., *F. chiloensis* and *F. virginiana*) of the cultivated strawberry (*Fragaria× ananassa*) (Govindarajulu et al. 2015; Njuguna et al. 2013; Tennessen et al. 2014). *Fragaria v.* subsp. *bracteata* is also notable because it is the only diploid *Fragaria* reported to have widespread sexual polymorphism (i.e., gynodioecy) (Ahmadi and Bringhurst 1989; Li et al. 2012). The male sterility reported in European *F. vesca* subsp. *vesca* (Irkaeva et al. 1993), in contrast, appears to be of limited occurrence (Ashman, T-L, unpublished data). This, plus the close affinity of *F. v.* subsp. *bracteata* to the octoploids, make it potentially influential in unraveling the evolution of male sterility and sex chromosomes in this clade (Tennessen et al. 2013). However, work-to-date has been based on only a few populations (Ahmadi and Bringhurst 1989; Li et al. 2012) and distinguishing the sexes has proven difficult

87 (Li et al. 2012). Thus, broader surveys of female frequencies (i.e., sex ratio), and evaluation of
88 male fertility of hermaphrodites (i.e., pollen viability) are needed to fully verify sexual system
89 and to place this variation in a geographic context.

90 Phylogenetic analysis suggests there may be a subdivision in *F. v.* subsp. *bracteata*
91 between the Pacific Coast and Rocky Mountain populations (Njuguna et al. 2013). Analysis of
92 the chloroplast genome places some samples in a clade that includes the sexually dimorphic
93 octoploid species while other samples group with *F. v*. subsp. *americana* and European *F. v*.
94 subsp. *vesca* (Govindarajulu et al. 2015; Njuguna et al. 2013). However, more populations are
95 needed to fully characterize the geographic pattern of chloroplast diversity and to evaluate it in
96 light of speculation concerning a history of hybridization in *F. v.* subsp. *bracteata* (Staudt 1999).
97 Finally, Staudt (1999) notes that one of the diploid taxa, *F. mexicana*, is often considered
98 synonymous with *F. v.* subsp. *bracteata*. However, the genetic affinity this taxon has relative to
99 all the subspecies has never been determined. Thus, documentation of the geographic patterns of
100 natural genetic variation for all the taxa in the North American diploid strawberry complex will
101 serve as an important point of reference for future ecological, evolutionary and functional
102 research in the strawberry genus (Liston et al. 2014; Shulaev et al. 2011).

103 We characterized the patterns of genetic variation in the four interfertile North American
104 diploid strawberries with special focus on *F. v.* subsp. *bracteata*. We verified the sexual system
105 differences of the taxa by characterizing female frequencies and pollen viability of
106 hermaphrodites. We determined whether sexual system affected nuclear genetic diversity across
107 all taxa. And finally, we used the genetic data to assess whether there has been a history of
108 hybridization within *F. v.* subsp. *bracteata* and to assess whether sexual system or sex ratio
109 affected population inbreeding level.

Materials and methods

Study System

The four diploid (2n = 2× = 14) *Fragaria* taxa have adjacent ranges in North America (Fig. 1). *Fragaria v.* subsp. *bracteata* and *F. v.* subsp. *americana* have the largest ranges, with *F. v.* subsp. *bracteata* extending from British Columbia south along the Pacific Coast into California and east into the Rocky Mountains and *F. v.* subsp. *americana* stretching from Canada's Northwest Territories west to British Columbia and south along the Atlantic Coast to Virginia (Staudt 1999). *Fragaria v.* subsp. *californica* is limited to coastal areas from southern Oregon to southern California with a disjunct population in southern Baja California, Mexico, whereas *F. mexicana* is widespread but disjunctly distributed at high elevations in Mexico south to the state of Chiapas (Staudt 1999). For the purposes of this study, we consider *F. mexicana* as distinct at the outset and evaluate its affiliations with *F. v.* subsp. *bracteata* directly via genetic markers.

All four taxa are spring-blooming herbaceous perennials. They reproduce sexually by seed (achenes) on fleshy "fruits" (swollen receptacle) or clonally by stolons. Three taxa (*F. v.* subsp. *americana*, *F. v.* subsp. *californica*, and *F. mexicana*) are hermaphroditic and self-compatible, whereas *F. vesca* subsp. *bracteata* is gynodioecious (Ahmadi and Bringhurst 1989; Li et al. 2012). Flowers of female and hermaphrodite *F. v.* subsp. *bracteata* are morphologically very similar, but anthers of pistillate flowers produce no viable pollen grains (Li et al. 2012). A nuclear locus that affects sex expression with a dominant allele for male sterility has been identified in material from a single population of *F. v.* subsp. *bracteata* from Oregon (Tennessen et al. 2013).

Sampling

To explore patterns of genetic variation across the subspecies and among nuclear and cytoplasmic genomes we used two sampling schemes: 1) shallow sampling (11 populations at one plant/population) and 2) deep sampling (15 populations at four to 18 plants/population). For the 11 shallowly sampled populations dried leaves were obtained from our collections (1) or from the National Clonal Germplasm Repository (10) (NCGR: http://www.ars.usda.gov/main/site_main.htm?modecode=53-58-15-00). For deep sampling, we sampled whole plants or fruits (one per maternal plant) along transects from nine populations of *F. v.* subsp. *bracteata*, two *F. v.* subsp. *californica*, three *F. v.* subsp. *americana* and one of *F. mexicana.* Plants or one seed per fruit were grown in the greenhouse at the University of Pittsburgh. For one deeply sampled population (ONT-GRL) leaves were obtained from field-collected seeds grown in the greenhouse by Brian Husband. Population names are represented as a combination of the state (province or country) of origin and a local place name (e.g., OR-MRD, 'Oregon' and 'Mary's Road'), GPS locations, collection sources and sampling intensity are summarized in Table A1.

DNA extraction

DNA was extracted from 100-150 mg of fresh leaf tissue or 20-30 mg of silica-dried leaf tissue per plant. We used a CTAB DNA protocol (Doyle and Doyle 1990), as modified in Salamone et al. (2013). Nuclear genomic DNA was quantified for PCR using a Spectromax 190 spectrophotometer (Molecular Devices, Sunnyvale, California, USA) and subsequently diluted with deionized sterile water (0.1-0.3ng/mL) as needed.

Amplification and sequencing of chloroplast and mitochondrial genes.

We amplified ca. 700 bp of the *rpoC2* chloroplast gene region using published primers (Table A2). The PCR reaction included 1X standard reaction buffer (New England Biolabs), 100 µM of each dNTP, 0.5 µM of each forward and reverse primers, 1.5 units of Standard *Taq* polymerase (New England Biolabs) and 1µl of genomic DNA in a 20 µl reaction. The amplification conditions were 2.5 minutes at 95°C, followed by 35 cycles of 95°C for 30 seconds, 52°C for 30 seconds, and 72°C for 60 seconds, and a final extension at 72°C for eight minutes.

To assess mitotype diversity, we sequenced 425-746 bp portions of ten mitochondrial genes (*atp4, atp6, atp8, atp9, cob, cox3, nad6, nad9, rps1*) and a novel ORF (*atp8-orf225*) using published primers or those developed herein (i.e., for *atp8, rps1* and *atp8-orf225*) (Table A2). PCR amplification used the conditions described above for the chloroplast gene with annealing temperatures between 48-54°C (Table A2).

The ten mitochondrial genes were screened in one individual from 10 populations to first test for polymorphism. Of these only two showed any polymorphism (*atp8* and *rps1*) and we selected the one with a greater number of mitotypes for further study (351 bp of *atp8*). The *atp8-orf225* is a novel 294 bp open reading frame (ORF) with a putative role in cytoplasmic male sterility and a phylogenetic signal inconsistent with the rest of the mitochondrial genome but consistent with the chloroplast genome (Govindarajulu et al. 2015; Tennessen et al. 2013). We amplified 425 bp of *atp8-orf225* (including flanking sequence to avoid co-amplification of the full length *atp8* gene) using primers in Table A2.

The amplified PCR products of both chloroplast and mitochondrial genes were purified using Exo-SAP (Affymetrix/USB, Cleveland, OH) and sequenced on the ABI 3730XL DNA Analyzer (Applied Biosystems, Sunnyvale, CA). Sequences were aligned using CodonCode Aligner (CodonCode Corporation, Centerville, MA) and edited using Bioedit (Hall 1999).

178 ***Nuclear genotyping***

179 For the 15 deeply sampled populations (Table A1) we also genotyped four to18 plants
180 (174 plants in total) at five nuclear microsatellite markers: CX661603, CX661264, UDF002,
181 ARSFL 7, and Fvi11 following Li et al. (2012). Each marker was amplified using the Poor
182 Man's protocol (Govindarajulu et al. 2013; Schuelke 2000). The PCR products for all five
183 primers were either run separately or multiplexed using 1.5 µl aliquots from each reaction with
184 0.2 µl LIZ500 standard and 10.5 µl Hi-Di formamide (Applied Biosystems, Carlsbad, CA).
185 Fragment analysis was performed using the ABI 3730XL DNA Analyzer. Product sizes were
186 visualized and genotyped using the program GeneMapper ver. 3.0 (Applied Biosystems).

187 ***Genetic analyses***

188 Sequences derived from all samples of *rpoC2* were aligned to visualize SNPs at two
189 informative positions 2934 bp (G and C) and 3045bp (T and C). Previously SNPs at the second
190 site (only) in this region were used to differentiate Pacific Coast *F. v.* subsp. *bracteata* from
191 Rocky Mountain *F. v.* subsp. *bracteata* and *F. v.* subsp. *americana* (Njuguna et al. 2013). Here
192 we present both SNP positions to facilitate comparison with past data sets, and refer to the G/T
193 and C/C as chlorotype 1 and 2, respectively. The nucleotide polymorphisms observed in *atp8* and
194 *atp8-orf225* were scored for each sequence and both separate and concatenated mitotype
195 designations were made for all samples. We calculated the number of mitotypes, number of
196 segregating sites and mitotype diversity within and across the deeply sampled populations using
197 DNaSP ver. 5.10.1 (Rozas et al. 2010). Additionally, we tested for evidence of recombination
198 between *atp8* and *atp8-orf225* in all individuals from the deeply sampled populations using the
199 four-gamete test (Hudson and Kaplan 1985) as implemented in DNaSP. Using all distinct
200 haplotypes found in the study (regardless of source population), we created individual minimum

spanning trees for *atp8, atp8-orf225*, concatenated *atp8-orf225+atp8* and *rpoC2* in Arlequin ver. 3.5.1.3 (Excoffier and Lischer 2010). We did not concatenate chloroplast and mitochondrial genes because there is evidence that they are unlinked (Govindarajulu et al. 2015). To visualize minimum spanning networks, we constructed haplotype networks in HapStar (Teacher and Griffiths 2011).

To characterize nuclear genetic diversity in the deeply sampled populations, we calculated the number of polymorphic loci in Microsatellite Toolkit ver. 3.1.1 (Park 2001). Allelic richness and Weir and Cockerham's global F_{ST} were estimated in FSTAT ver. 2.9.3.2 (Goudet 2002). F_{ST} was calculated for all taxa together and then for *F. v.* subsp. *bracteata* alone. Wright's inbreeding coefficient (F_{IS}) was only calculated for populations with more than one polymorphic locus, and only plants with genotypic data for at least three of the five SSR loci were included in these analyses (n = 166). Allelic richness was used as an indicator of population admixture and size because it can detect bottlenecks even before substantial loss of heterozygosity occurs (Comps et al. 2001; Luikart et al. 1998).

To investigate admixture between the subspecies, we performed a Bayesian model-based clustering analysis in STRUCTURE ver. 2.3.3 (Pritchard et al. 2000) on the 15 deeply sampled populations. Individuals (n = 166) were assigned probabilistically to *K* inferred genetic clusters, each characterized by a set of allele frequencies at Hardy-Weinberg equilibrium and in linkage equilibrium. We ran the program with the admixture model and independent allele frequencies for a range of *K* values from 1 to 15. We used 10,000 burn-ins and Monte Carlo Markov Chain replicates for each run, with 10 replicate runs conducted for each *K* value. We used Structure Harvester (Earl and vonHoldt 2012) to calculate a ΔK value that indicates the optimal number of genetic

clusters based on methods from Evanno et al. (2005). The final matrix of nuclear membership for each individual and population was visualized using DISTRUCT ver. 1.1 (Rosenberg 2007).

The geographic distribution of the chloroplast (*rpoC2*) and mitochondrial (*atp8-orf225* and *atp8*) haplotypes was visualized in the program DIVA-GIS (Hijmans et al. 2007). Data from the *F. v.* subsp. *bracteata* individuals from nine locations with known *F. v.* subsp. *bracteata* or *F. v.* subsp. *americana*-type chloroplast SNPs (Njuguna et al. 2013) were also included on the map for comparative purposes.

Sexual system

For each deeply sampled population, the proportion of females was estimated from either surveys (sexual identity of flowering plants every 2 m [to avoid sampling the same clone] along two to three 20m transects in the field; as in Ashman 1999) or from sexual identity of greenhouse grown plants (or one seed per fruit) that were collected similarly along transects in the field. Sample size per population (Table 1) depended on plant abundance and ranged. Hermaphrodite plants had plump yellow anthers that visibly shed pollen, whereas individuals whose anthers were lighter yellow or pale and that never dehisced were categorized as female (following Li et al. 2012). We assessed pollen viability of hermaphrodite plants in the deeply sampled populations by collecting dehiscing anthers from two freshly opened flowers per plant grown under common greenhouse conditions. Pollen was stained with Alexander's stain and viability was scored visually under a light microscope for >100 pollen per flower (Kearns and Inouye 1993). The proportion of viable grains was calculated for each flower, averaged per plant, and the mean calculated for each population from plant means. Anthers from individuals classified as females were also assessed using this method to confirm qualitative assessment of sexual identity (i.e., no pollen produced). Populations with no females were considered hermaphroditic, while

247 populations with female individuals were considered gynodioecious.

248 ***Sexual system analyses***

249 To assess whether sexual system affected nuclear genetic diversity (population-
250 level inbreeding, allelic richness), we conducted a *t*-test comparing the mean nuclear
251 genetic diversity metrics of all populations from hermaphroditic taxa (*F. v.* subsp.
252 *americana*, *F. v.* subsp. *californica*, and *F. mexicana*) to that of populations from the
253 gynodioecious subspecies (*F. v.* subsp. *bracteata*). We also evaluated the relationship
254 between female frequency, pollen viability and nuclear genetic diversity across all *F. v.*
255 subsp. *bracteata* populations using Pearson correlations. Pearson correlations are viewed
256 as an adequate choice even for zero inflated data (Huson 2007). All analyses were
257 implemented in R (R Core Development Team 2012).

258 Nuclear microsatellite data were deposited at Dryad (http://dx.doi.
259 org/XXX), while chloroplast and mitochondrial gene sequences were deposited in Genbank *atp8*:
260 KM107197 - KM107311; *atp8-orf225*: KM107312 - KM107426; *rpoC2*: KM107427 -
261 KM107563).

262

263 **Results**

264 ***Diversity in chloroplast and mitochondrial genes***

265 Sequencing of chloroplast gene *rpoC2* in 25 wild populations of diploid *Fragaria*
266 in North America (Fig. 2*a*) confirmed the existence of two distinct types (Njuguna et al.
267 2013). All deeply sampled populations of *F. v.* subsp. *americana* were chlorotype 1, but
268 those of *F. v.* subsp. *bracteata* were either monomorphic for chlorotype 1 (four
269 populations) or 2 (three populations), or were polymorphic (three populations: % type 1:

OR-ES: 60%; OR-FB: 10%; ID-SFR: 78%). The deeply sampled *F. v.* subsp. *californica* populations were monomorphic for chlorotype 2, whereas *F. mexicana* was monomorphic for chlorotype 1, in agreement with the single shallowly sampled population of this subspecies (MX-PE). Thus, *F. v.* subsp. *californica* was indistinguishable from Pacific Coast *F. v.* subsp. *bracteata*, while *F. mexicana* was indistinguishable from Rocky Mountain *F. v.* subsp. *bracteata* and *F. v.* subsp. *americana* at the chloroplast gene *rpoC2*.

Both mitochondrial genes amplified in all but one population. A total of eight SNPs were found, three in *atp8* and five in *atp8-orf225*. We did not detect any recombination events between *atp8* and *atp8-orf225* based on our four-gamete test (Hudson and Kaplan 1985). Thus there were eight unique concatenated *atp8-orf225+atp8* sequences, referred to as mitotypes A-H (Table 1, Fig. 2*b*; Table A3). Minimum spanning networks generated from concatenated *atp8-orf225+atp8* sequence characterized mitotype H as the most genetically distant from mitotype A. Most *F. v.* subsp *bracteata*, both *F. v.* subsp. *californica*, and *F. mexicana* populations contained mitotype B or C, and almost all of them (11/12) had the same SNP combination in *atp8* (See Table A3). The exception was population NM-LNF, which was fixed for a novel mitotype (F). In contrast, the *F. v.* subsp. *americana* populations (NH-FP and ONT-GRL) had mitotype H, or were missing concatenated mitotype information (IA-SSP) because the *atp8-orf225* would not amplify [although *atp8* would]. Population ONT-GRL also had two individuals with mitotype D. Deep sampling revealed polymorphism within several populations of *F. v* subsp. *bracteata*: three were dimorphic (OR-MRD C/B: 90/10%; ID-SFR E/C:70/30% and CO-SCR C/G:70/30%) and one was trimorphic (OR-ES- G/C/E: 50/40/10%). Mitotype diversity in these populations ranged from 0.20-0.64 (Table 1). None of the deeply sampled *F. v.* subsp. *californica*, *F. mexicana*, nor two of the three *F. v.* subsp. *americana* populations showed mitotypic polymorphism (Table 1).

293 The single population of *F. v.* subsp. *americana* with polymorphism (ONT-GRL) was
294 unique among the polymorphic populations in that the two individuals with minority
295 mitotypes also had divergent nuclear affiliations (see below). It is also noteworthy that
296 with the exception of the two individuals with mitotype D in ONT-GRL, the rarer
297 mitotypes (E, F, and G) generally occur in the center of the network, showing sequences
298 intermediate between the more common B/C mitotypes and the H mitotype.
299 Networks based on *atp8* and *atp8-orf225* sequences separately revealed four
300 distinct mitotypes. Almost all *F. v.* subsp. *americana* individuals shared the GAGGC
301 *atp8* mitotype, whereas the western subspecies shared three other mitotypes (Fig. 2*b*;
302 Table A3). For *atp8-orf225* there was a common single nucleotide polymorphic position
303 (TCG) that most populations contained, and three divergent sequences, one shared by
304 four *F. v* subsp. *bracteata* populations (two in Oregon [OR-MRD, OR-FB] and two in
305 California [CA-WMR, CA-PDL]), one unique to the high elevation *F. v* subsp. *bracteata*
306 population in New Mexico (NM-LNF), and one unique to two individuals of the *F. v.*
307 subsp. *americana* in Guelph Canada (ONT-GRL). Mitotype diversity was high in these
308 polymorphic populations (Table 1).
309 ***Geographic distribution of cytotypes reveals zone of admixture and novelty***
310 Geographic mapping of both deeply and shallowly sampled populations in
311 combination with data from Njuguna et al. (2013) revealed an east-west distribution of
312 chlorotypes within *F. v.* subsp. *bracteata* (Fig. 3). Populations located west of the
313 Rockies were dominated by chlorotype 2 while those to the east were dominated by
314 chlorotype 1, thus following the distributional ranges of the widespread subspecies (*F. v.*
315 subsp. *bracteata* and *F. v.* subsp. a*mericana*). A hybrid zone extending from the Oregon

316 Cascades to western Idaho is suggested by the three *F. v.* subsp. *bracteata* populations (OR-ES,
317 OR-FB, and ID-SFR) with mixed chlorotypes.

318 Although the geographic distribution of mitotypes is less straightforward than for the
319 chlorotypes, a longitudinal pattern was seen: The westernmost populations showed mitotypes B
320 and/or C (all *F. v.* subsp. *californica*, *F. mexicana* and many *F. v.* subsp. *bracteata)*, the
321 easternmost populations (all *F. v.* subsp. *americana*) had mitotype H, whereas, the more central
322 populations of *F. v.* subsp. *bracteata* showed admixture plus the occurrence of novel mitotypes
323 (E, F, and G). Two of the populations (OR-ES, ID-SFR) that were polymorphic in the
324 mitochondrial genome also showed polymorphism in the chloroplast. These populations are
325 between the Oregon Cascade and the Rocky Mountains in Idaho, and show the highest degree of
326 cytoplasmic admixture (between them, three of the eight observed mitotypes and both of the
327 chlorotypes are represented). Separate from these geographic patterns, a novel mitotype (D) was
328 found in two individuals of ONT-GRL amongst other individuals with the majority mitotype (H)
329 for *F. v.* subsp. *americana* (Figs. 2,3).

330 ***Nuclear genetic variation shows differentiation and distinct population structure***

331 Analysis of microsatellite variation in the 15 deeply sampled populations revealed one to
332 five polymorphic loci per population and 1.1 to 2.6 alleles per locus (Table 1). There was
333 significant genetic differentiation among all North American diploid strawberry populations (F_{ST}
334 = 0.588) and among the nine *F. v.* subsp. *bracteata* populations (F_{ST} = 0.476). Inbreeding ranged
335 10-fold (F_{IS} = 0.065 to 0.760; Table 1) with significant inbreeding in six populations (4 of the 9
336 *F. v.* subsp. *bracteata,* one population of *F. v.* subsp. *californica,* and *F. mexicana*; Table 1). One
337 *F. v.* subsp. *californica* population (CA-SBNF) was only polymorphic at one locus, potentially
338 indicating high inbreeding as well.

Structure analysis indicated the presence of two distinct genetic groups ($\Delta K = 2$; –lnL = -1971.12; Fig. 4). The three populations of *F. v.* subsp. *americana* showed a high proportion of nuclear membership to cluster 1 (92 ± 1%), whereas eight of the nine populations of *F. v.* subsp. *bracteata* showed a high proportion of nuclear membership to cluster 2 (97 ± 2%). The only population to deviate from this pattern was NM-LNF, which showed high membership to cluster 1 (97%). The southern *F. v.* subsp. *californica* and *F. mexicana* both showed affinity with cluster 2 (99 ± 0.05%). Overall, the geographic pattern of nuclear genetic variation followed the same east-west distribution seen in the chloroplast, which is consistent with the known ranges of *F. v.* subsp. *americana* and *F. v.* subsp. *bracteata*. There was little nuclear admixture observed within populations (less than 5 %) despite the admixture seen in the cytogenomes of three *F. v.* subsp. *bracteata* populations (OR-ES, OR-FB, and ID-SFR). Plants within these three populations all had nuclear affiliations to cluster 2 but a percentage of individuals within each had chlorotype 1 (OR-ES: 60%, OR-FB: 10%, ID-SFR: 78%). We also detected two populations for which all sampled individuals demonstrated a mismatch between the nuclear and chloroplast affiliations: CO-SCR and MX-LV are strongly affiliated with cluster 2 (*F. v.* subsp. *bracteata* nuclear genome) but are both monomorphic for chlorotype 1 (*F. v.* subsp. *americana* type). The northern populations are located along the Rocky Mountains where hybridization is likely to occur. The other evidence of admixture comes from ONT-GRL, a population well within *F. v.* subsp. *americana* range, but hosting a minority of individuals (2/10) that cluster with the *F. v.* subsp. *bracteata* (Fig. 4). These individuals are also novel in mitotype and could reflect recent migrants (potentially human introduced) or early generation hybrids.

Sex expression and correlates with population genetic diversity

Consistent with current understanding of sexual systems in the North American diploid strawberries, females were only found in *F. v.* subsp. *bracteata* populations where they ranged in frequency from 0 - 46% (Table 1). Moreover, hermaphrodites of *F. v.* subsp. *bracteata* had significantly lower mean proportion pollen viability than those of the hermaphroditic taxa (*F. v.* subsp. *americana*, *F. v.* subsp. *californica* and *F. mexicana*; Table 1) (mean ± s.e.: 0.77 ± 0.02 *vs.* 0.91 ± 0.02, t_{13} = -4.70, p = 0.001).While the inbreeding coefficient was not significantly different (*F. v.* subsp. *bracteata*: 0.30 ± 0.07 *vs.* hermaphroditic taxa: 0.49 ± 0.13, t_{13} = -1.32, p = 0.21), nuclear allelic richness was higher in populations of gynodioecious *F. v.* subsp. *bracteata* than those of the hermaphroditic taxa (*F. v.* subsp. *americana*, *F. v.* subsp. *californica* and *F. mexicana*) (*F. v.* subsp. *bracteata*: 2.04 ± 0.12 *vs*. hermaphroditic taxa: 1.60 ± 0.13, t_{13} = 2.43, p = 0.03; Table 1). Among the 9 populations of gynodioecious *F. v.* subsp. *bracteata*, however, we did not detect a significant correlation of female frequency with allelic richness, inbreeding nor mean pollen viability (all $p > 0.10$). Although we note that some sample sizes are small.

Discussion

Geographic patterns of genetic variation in North American diploid strawberries

Geographic analysis of the three genomes of North American diploid *Fragaria* confirmed the known distributions of the widespread subspecies and gave insight into genetic affiliations of the more restricted taxa. Specifically, the structure analysis of nuclear markers aligned *F. v.* subsp. *californica* and *F. mexicana* with *F. v.* subsp. *bracteata*. Thus, these genetic results confirm Staudt's (1999) lumping of *F. mexicana* with *F. v.* subsp. *bracteata* based on phenotypic data. These two rarer taxa also reflect the chlorotype division observed in the wide spread *F. v.*

subsp. *bracteata* with *F. v.* subsp. *californica* possessing the Pacific Coast chlorotype and *F. mexicana* the Rocky Mountain chlorotype.

The mitochondria showed similar broad scale patterns to the chloroplast (especially with respect to *F. v.* subsp. *californica* and *F. mexicana*), but revealed the potential for finer scale diversity in the widespread species as well. Many populations were polymorphic and some showed entirely novel mitotypes. For instance, within *F. v.* subsp. *americana* a novel mitotype was found in the Guelph, Ontario population along with two individuals that were potentially recent immigrants (i.e., were in the minority but belonged to the *F. v.* subsp. *bracteata* nuclear cluster). Unfortunately, human introduction cannot be ruled out, considering the proximity to University of Guelph, where diverse *Fragaria* species have been grown in the past (Bors and Sullivan 2005). Whether this reflects the introduction of a novel mitotype in a *F. v.* subsp. *bracteata* migrant, or if it is the result of undetected intragenomic recombination facilitated by heteroplasmy (McCauley 2013) in association with hybridization (Kmiec et al. 2006) is unknown. Additionally, in another *F. v.* subsp. *americana* population (IA-SSP) the primers for *atp8-orf225* did not amplify a product. This could reflect the loss of this gene from the mitochondria in this population (as in Kmiec et al. 2006) or simply a mutation in the primer site. Nonetheless, these data suggest that these mitochondrial genes are valuable for tracking fine-scale geographic patterns. In fact, if *atp8-orf225* (Govindarajulu et al. 2015), is confirmed to have had a functional role in male sterility, then these patterns could reflect past cyto-nuclear evolutionary dynamics (e.g., Städler and Delph 2002) in a species that at present is phenotypically hermaphroditic.

Geographic patterns of genetic variation reveal hybridization in F. v.* subsp. *bracteata

406 Our landscape-level analysis also identifies a zone of cytoplasmic introgression between
407 the Cascades and Northern Rocky Mountains where the ranges of *F. v.* subsp. *bracteata* and *F. v.*
408 subsp. *americana* contact. This discovery of admixed populations west of the Rocky Mountains
409 is consistent with Staudt's (1999) observations of hybrids between *F. v.* subsp. *americana* and *F.*
410 *v.* subsp. *bracteata* in Yellowstone National Park and southwest of the Rocky Mountains. These
411 data plus Staudt's (1999) reports of hybrids between *F. v.* subsp. *bracteata* and *F. v.* subsp.
412 *californica* in Curry county in southern Oregon and Humboldt and Del Norte counties in
413 northern California define a hotspot of past admixture between these subspecies in Northern
414 California/Oregon/Idaho. This region is also an active area of hybridization in the octoploid
415 strawberries (Salomone et al. 2013) suggesting that both diploids and octoploids may have
416 responded in parallel to past climate change. In contrast with the octoploid hybrid populations,
417 however, *F. v.* subsp. *bracteata* populations showed few admixed individuals, consistent with the
418 high degree of population structure observed in the nuclear genome across this large area. Rocky
419 Mountain *F. v.* subsp. *bracteata* display a strong signal of introgression of the *F. v.* subsp.
420 *americana* chloroplast into *F. v.* subsp. *bracteata* nuclear background, suggesting long distance
421 seed dispersal followed by pollen swamping (Beatty et al. 2010), or extensive backcrossing after
422 a past hybridization event, as in *Vasconcellea* (Van Droogenbroeck et al. 2006) and *Eucalyptus*
423 (McKinnon et al. 1999). Alternatively, it could reflect chloroplast capture, possibly facilitated by
424 gynodioecy in this subspecies. Specifically, Tsitrone and colleagues (2003) show theoretically
425 that a combination of female seed advantage and partial selfing can lead to rapid replacement of
426 chloroplasts in hybridizing lineages. In the context of *Fragaria,* Li et al. (2012) showed that both
427 of these predicted promoters of chloroplast capture exist within *F. v.* subsp. *bracteata*.
428 Moreover, Govindarajulu et al. (2015) has ruled out incomplete lineage sorting for the taxa in

question, and there is high support that *F. v*. subsp. *americana* and *F. v.* subsp. *bracteata* are in distinct clades, with *F. mexicana* and *F. v.* subsp. *californica* being in the same clade as *F. v.* subsp. *bracteata* (Syring et al. ms. in prep.). Thus, we conclude that the pattern of chloroplast-nuclear mismatches documented here reflect recent signatures of introgression where the taxa come in contact.

It is interesting to note as well that two distant populations in Colorado and Mexico (CO-SCR, MX-LV) have the same cyto-nuclear incongruence (*F. v*. subsp. *americana* chloroplast with *F. v.* subsp. *bracteata* nuclear genome) and one could speculate that it reflects range expansion of a single hybrid lineage, or strong ecological advantage of this combination over the alternative (*F. v*. subsp. *bracteata* chloroplast with *F. v.* subsp. *americana* nuclear genome). The adaptive value of sequence variation of cytoplasmic genomes, as well as the opportunity for ecological differentiation by nuclear-cytoplasmic interactions is only just now being appreciated (reviewed in Bock et al. 2014). Specific adaptive effects of cytoplasmic genome variation have been observed with regard to drought and thermal tolerances (Ballard and Melvin 2010; Sambatti et al. 2008), which could also be important for diploid *Fragaria* along the continental divide. The NM-LNF population is also interesting as it is a high elevation 'island' in the midst of desert and thus could be a relic of past glaciation, or simply genetic drift. This population has a high frequency of females, which could also be the result of founder events and restricted gene flow. These types of forces have been found to be the primary determinants of sex ratio in other gynodioecious species (e.g., *Beta vulgaris;* DeCauwer et al. 2012). More populations along this southern range are needed to understand the patterns of genetic admixture and phylogeography of diploid *Fragaria* in this region and to test these hypotheses explicitly. Artificial hybridization and transplant experiments would also be particularly useful (e.g., Sambatti et al. 2008).

***Sexual system variation and its effects on genetic diversity in* F. v. *subsp.* bracteata**

Eight of the nine *F. v.* subsp. *bracteata* populations showed clear evidence of females confirming the existence of gynodioecy in this subspecies over a much wider geographic range than previously studied (Ahmadi and Bringhust 1989). Sex ratio varied dramatically (from 0-46% females) as is common in gynodioecious species (e.g., Medrano et al. 2005; Nilsson and Ågren 2006), but there was no relation between female frequency and population-level inbreeding. Although such a relation has been found in three species (*Kallstroemia grandiflora* [Cuevas et al. 2006], *Daphne laureola* [Medrano et al 2005], and *Schizopepon bryoniaefolius* [Akimoto et al. 1999]), our work joins the findings of Tarayre and Thomspson (1997) who failed to find a negative relation between female frequency and population-level inbreeding in *Thymus vulgaris.* We hypothesize that factors such as gene flow or drift, as the result of past hybridization or population bottlenecks, have stronger fingerprints on genetic diversity than gynodioecy, or that biparental inbreeding occurs in females neutralizing their expected effect on population inbreeding level. The latter explanation may hold for *Thymus vulgaris* where there was no significant difference in inbreeding between females and hermaphrodites (Tarayre and Thompson 1997). In contrast, however, in *F. v.* subsp. *bracteata,* Li et al. (2012) found no biparental inbreeding by females but significant inbreeding by hermaphrodites. So if the results in one population (OR-MRD) studied by Li et al. (2012) are representative then the explanation may instead lie with the patterns of gene flow and drift. The present work provides evidence of the existence of the former: we found introgression in about half of the *F. v*. subsp. *bracteata* populations. A final alternative is that our sample size of populations or individuals was not adequate to reveal a relationship that exists. Deeper and broader sampling and knowledge of population size would ultimately be needed to distinguish among these possibilities.

Interestingly, the lower average pollen viability for hermaphrodites from *F. v.* subsp. *bracteata* populations compared to those from hermaphroditic taxa could also reflect hybridization because frequently hybridizing species can have lower pollen viabilities than rarely hybridizing ones (e.g., *Cirsium,* Bureš et al. 2010). Specifically, cyto-nuclear mismatches can cause pollen dysfunction, as has been seen in *Helianthus* and *Oenothera* (Levin 2003; Stubbe and Steiner 1999), and are thought to result from breakdown of coadapted interactions between organelle and nuclear genomes (reviewed in Burton et al. 2013). Alternatively, pollen inviability may be directly related to the influence of the chimeric open reading frame *atp8-orf225*, which is known to be expressed (Liston A. and R.C. Cronn, unpublished data), and is a candidate for a cytoplasmic male sterility gene (Govindarajulu et al. 2015). If the latter is the case, then populations must vary in the degree that expression is regulated by other factors (e.g., Dufay et al. 2008).

Conclusions and future work

This landscape-level study of genetic variation in the four diploid strawberry taxa adds important genetic insight to the biogeographic history of North American strawberries. It confirms the differences in sexual system among the taxa and starts to untangle the effects on genetic diversity. Future work in *F. v.* subsp. *bracteata* should be aimed at expanding the sampling in the southern portion of its range and determining the bioclimatic triggers of the geographic pattern of cyto-nuclear association observed in its nothern hybrid zone.

Acknowledgements

The authors thank A. Grant, G. Arceo-Gomez, G. Meindl, S. Falk, M. Koski, M. Parks, T. Rabogliatti, K. Schuller for greenhouse, field or laboratory assistance, the Ashman lab for discussion, B. Husband for leaf tissue from plants from ONT-GRL, the editor and two anonymous reviewers for comments that improved the manuscript. This work was supported by the University of Pittsburgh, the National Science Foundation (DEB 1020523 and RET/REU supplements to TLA and DEB 1020271 to AL), and HHMI Fellowship to LS.

References

Ahmadi, H., and Bringhurst, R.S. 1989. Genetics of sex expression in *Fragaria* species. Am. J. Bot. **78**:504-514.

Akimoto, J., Fukuhara, T., and Kikuzawa, K. 1999. Sex ratios and genetic variation in a functionally androdioecious species, *Schizopepon bryoniaefolius* (Cucurbitaceae). Am. J. Bot. **86**: 880–886.

Ashman, T-L. 1999. Determinants of sex allocation in a gynodioecious wild strawberry: implications for the evolution of dioecy and sexual dimorphism. J. Evol. Biol. **12**:648-661.

Ballard, J., and Melvin, R. 2010. Linking the mitochondrial genotype to the organismal phenotype. Mol. Ecol. **19**:1523-1539.

Barrett, S.C.H. 2003. Mating strategies in flowering plants: the outcrossing–selfing paradigm and beyond. Phil. Trans. R. Soc. Lond. **358**:991-1004.

Beatty, G.E., Philipp, M., and Provan, J. 2010. Unidirectional hybridization at a species' range boundary: implications for habitat. tracking. Div. Distr. **16**:1-9.

519 Bock, D., Andrew, R., and Rieseberg, L. 2014 On the adaptive value of cytoplasmic genomes in
520 plants. Mol. Ecol. **23**:4899-4911.
521 Bors, R., and Sullivan, J. 2005. Interspecific hybridization of *Fragaria vesca* subspecies with *F.*
522 *nilgerrensis*, *F. nubicola*, *F. pentaphylla*, and *F. viridis*. J. Amer. Soc. Hort. Sci. **130**:418-
523 423.
524 Bureš, P., Šmarda, P., Rotreklová, O., Oberreiter, M. Burešová, M., Konečný, J., Knoll, A.,
525 Fajmon, K., and Šmerda, J. 2010. Pollen viability and natural hybridization of Central
526 European species of *Cirsium*. Preslia **82**:391-422.
527 Burton, R.S., Pereira, R.J., and Baretto, F.S.. 2013, Cytonuclear genomic interactions and hybrid
528 breakdown. Annu. Rev. Ecol. Evol. Syst. **44**:281–302
529 Comps, B., Gömöry, D., Letouzey, J., Thiébaut, B., and Petit, R.J. 2001. Diverging trends
530 between heterozygosity and allelic richness during postglacial colonization in the
531 European Beech. Genetics. **157**:389-397.
532 Cuevas, E., Arias, D.M., Dominguez, C.A., Castillo, R.A., and Molina-Freaner, F. 2006. The
533 genetic structure of the gynodioecious *Kallstroemia grandiflora* (Zygophyllaceae): the
534 role of male sterility and colonization history. Heredity **97**: 269–274.
535 Dalton, R., Koski, M., and Ashman, T-L. 2013. Maternal sex effects and inbreeding depression
536 under varied environmental conditions in gynodioecious *Fragaria vesca* subsp. *bracteata*
537 Ann. Bot. **112**:613-621.
538 DeCauwer, L., Dufay, M., Hornoy, B., Courseaux, A., and Arnaud, J-F. 2012. Gynodioecy in
539 structured populations: understanding fine-scale sex ratio variation in *Beta vulgaris* ssp.
540 *maritime.* Mol. Ecol. **21**:834-850.
541 Doyle, J.J., and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. Focus **12**:13-15.

542 Dufay, M., Vaudey, V., DeCauwer, I., Touzet, P., Cuguen, J., Arnaud, J. 2008. Variation in
543 pollen production and pollen viability in natural populations of gynodioecious *Beta*
544 *vulgaris* ssp. *maritima*: evidence for a cost of restoration of male function? J. Evol. Biol.
545 **21**: 202–212.
546 Earl, D.A., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for
547 visualizing STRUCTURE output and implementing the Evanno method. Cons. Gen. Res.
548 **4**:359-361.
549 Eckert, C.G., Samis, E., and Lougheed, S.C. 2008. Genetic variation across species'
550 geographical ranges: the central–marginal hypothesis and beyond. Mol. Ecol **17**:1170-
551 1188.
552 Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals
553 using the software structure: a simulation study. Mol. Ecol. **14**:2611-2620.
554 Excoffier, L., and Lischer, H.E.L. 2010, Arlequin suite ver 3.5: A new series of programs to
555 perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. **10**:564-
556 567.
557 Gitzendanner. M.A. and Soltis, P.S. 2000. Patterns of genetic variation in rare and widespread
558 plant congeners. Am. J. Bot. **87**:783-792.
559 Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices.
560 version 2.9.3.2.
561 Govindarajulu, R., Liston, A., and Ashman, T-L. 2013. Sex-determining chromosomes and
562 sexual dimorphism: insights from genetic mapping of sex expression in a natural hybrid
563 *Fragaria* ×*ananassa* subsp. *cuneifolia*. Heredity **110**:430-438.

564 Govindarajulu, R., Parks, M., Tennessen, J.A., Liston, A., and Ashman, T-L. 2015. Comparison
565 of nuclear, plastid and mitochondrial phylogenies and the origin of wild octoploid
566 strawberry species. Am. J. Bot. **102**: 544-554.
567 Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis
568 program for Windows 95/98/NT. Nucl. Acid. Symp. Series. **41**:95-98.
569 Hamrick, J.L. and Godt, M.J.W. 1989. Allozyme diversity in plant species. *In* A. H. D. Brown,
570 M. T. Clegg, A. L. Kahler, and B. S. Weir [eds.], Plant population genetics, breeding, and
571 genetic resources pp. 43–63. Sinauer, Sunderland, Massachusetts, USA.
572 Hijmans, R.J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D.M. 2007.
573 Geographical and environmental range expansion through polyploidy in wild potatoes
574 (*Solanum* section Petota). Global Ecol. Bio. **16**:485-495.
575 Hudson, R.R., and Kaplan, N. 1985. Statistical properties of the number of recombination events
576 in the history of a sample of DNA sequences. Genetics **111**:147–165.
577 Huson, L.W. 2007. Performance of some correlation coefficients when applied to zero-clustered
578 data. J of Mod Appl Stat Meth. **6**:530-536.
579 Irkaeva, N.M., Lebedeva, L.V., and Shaziyatova, D.S. H. 1993. Genetics of male sterility in the
580 strawberry *Fragaria vesca* L. Soviet Genetics **29**:1103-1107.
581 Kearns, C.A., and Inouye, D.W. 1993. Techniques for Pollination Biologists. Univ. Press of
582 Colorado, Boulder.
583 Kmiec, B., Woloszynska, M., and Janska, H. 2006. Heteroplasmy as a common state of
584 mitochondrial genetic information in plants and animals. Curr. Genet. **50**:149-159.
585 Levin, D.A. 2003. The cytoplasmic factor in plant speciation. Syst. Bot. **28**:5–11

586 Li, J., Koski, M.H., and Ashman, T-L. 2012. Functional characterization of gynodioecy in
587 *Fragaria vesca* ssp. *bracteata* (Rosaceae). Ann. Bot. **109**:545-552.
588 Liston, A., Cronn, R.C., and Ashman, T-L. 2014. Fragaria: a genus with deep historical roots and
589 ripe for evolutionary and ecological insights. Am. J. Bot. **101**:1686-1699.
590 Luikart, G., Sherwin,W.B., Steele, B.M., and Allendorf, F.W. 1998. Usefulness of molecular
591 markers for detecting population bottlenecks via monitoring genetic change. Mol. Ecol.
592 **7**:963-974.
593 Manel, S. Schwartz, M.K., Luikart, G. and Taberlet. P. 2003. Landscape genetics: combining
594 landscape ecology and population genetics. Trends Ecol. Evo. **18**:189-197.
595 Matosiuk, M., Borkowska, A., Świsłocka, M., Mirski, P., Borowski, Z., Krysiuk, K., Danilkin,
596 A.A., Zvychaynaya, E.Y., Saveljev, A.P., and Ratkiewicz, M. 2014. Unexpected
597 population genetic structure of European roe deer in Poland: an invasion of the mtDNA
598 genome from Siberian roe deer. Mol. Ecol. **23**:2559-2572.
599 McCauley, D. 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial
600 genomes. New Phytol. **200**:966-977.
601 McKinnon, G., Steane, D., Potts, B., and Vaillancourt, R. 1999. Incongruence between
602 chloroplast and species phylogenies in *Eucalyptus* subgenus Monocalyptus (Myrtaceae).
603 Am. J. Bot. **86**:1038-1046.
604 Medrano, M., Alonso, C., and Herrera, C.M. 2005. Mating system, sex ratio, and persistence of
605 females in the gynodioecious shrub *Daphne laureola* L. (Thymelaeaceae). Heredity **94**:
606 37–43.
607 Nilsson, E. and Ågren. J. 2006. Population size, female fecundity, and sex ratio variation in
608 gynodioecious *Plantago maritime.* J. Evol. Biol. **19**:825–833.

609 Njuguna, W., Liston, A., Cronn, R.C., Ashman, T-L., and Bassil, N. 2013. Insights into
610 phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome
611 sequencing. Mol. Phylo. Evol. **66**:17-29.
612 Park, S. 2001 Trypanotolerance in West African cattle and the population genetic effects of
613 selection. PhD diss. University of Dublin, Ireland.
614 Pritchard, J., Stephens, M. and Donnelly, P. 2000. Inference of population structure using
615 multilocus genotype data. Genetics **155**:945-959.
616 R Core Development Team 2012 R statistical software.
617 Rosenberg, N.A. 2007. Distruct: a program for the graphical display of population structure.
618 https://web.stanford.edu/group/rosenberglab/distruct.html.
619 Rozas, J., Librado, P., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. 2010. DNASSP
620 version 5.10.1.http://www.ub.edu/dnasp/.
621 Salamone, I., Govindarajulu, R., Falk, S., Parks, M., Liston, A., and Ashman, T-L. 2013
622 Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a
623 natural hybrid of octoploid strawberries. Am. J. Bot. **100**:939-950.
624 Sambatti, J., Ortiz-Barrientos, D., Baack, E., and Rieseberg, L. 2008. Ecological selection
625 maintains cytonuclear incompatibilities in hybridizing sunflowers. Ecol. Let. **11**:1082-
626 1091.
627 Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature
628 Biotech. **18**:233-234.
629 Seneviratne, S., Toews, D.P.L., Brelsford, A. and Irwin, D.E. 2012. Concordance of genetic and
630 phenotypic characters across a sapsucker hybrid zone. J. Avian Biol. **43**:119-130.

631 Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal,
632 P., Mockaitis, K., Liston, A., Mane, S.P., Burns, P., Davis, T.M., Slovin, J.P., Bassil, N.,
633 Hellens, R.P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O.R., Jensen, R.V.,
634 Allan, A.C., Michael, T.P., Setubal, J.C., Celton, J-M., Rees, D.J.G., Williams, K.P.,
635 Holt, S.H., Ruiz Rojas, J.J., Chatterjee, M., Liu, B., Silva, H. Meisel, L., Adato, A.
636 Filichkin, S., Troggio, M., Viola, R., Ashman, T-L., Wang, H., Dharmawardhana, P.,
637 Elser, J., Raja, R., Priest, H.D., Bryant Jr., D.W., Fox, S.E., Givan, S.A., Wilhelm, L.J.,
638 Naithani, S., Christoffels, A., Salama, D.Y., Carter, J., Lopez Girona, E., Zdepski, A.,
639 Wang, W., Kerstetter, R.A., Schwab, W., Korban, S.S., Davik, J., Monfort, A., Denoyes-
640 Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J.L., Salzberg, S.L.,
641 Dickerman, A.W., Velasco, R., Borodovsky, M., Veilleux, R.E. and Folta, K.M. 2011.
642 The genome of woodland strawberry, *Fragaria vesca*. Nature Gen. **43**:109–116
643 Soltis, P.S., and Soltis, D.E. 2009. The role of hybridization in plant speciation. Ann. Rev. Plant
644 Biol. **60**:561-588.
645 Städler, T., and Delph, L. 2002. Ancient mitochondrial haplotypes and evidence for intragenic
646 recombination in a gynodioecious plant. Proc. Nat. Acad. Sci. **99**:11730-11735.
647 Staudt, G. 1999. Systematics and geographic distribution of the American strawberry species :
648 taxonomic studies in the genus *Fragaria* (Rosaceae:Potentilleae). University of
649 California Press, Berkeley.
650 Stubbe, W., and Steiner, E. 1999. Inactivation of pollen and other effects of genome-plastome
651 incompatibility in *Oenothera*. Plant Syst. Evol. **217**:259–77.
652 Tarayre, M., and Thompson, J.D. 1997. Population genetic structure of the gynodioecious
653 *Thymus vulgaris* L. (Labiatae) in southern France. J. Evol. Biol. **10**:157-174.

654 Teacher, A.G.F., and Griffiths, D.J. 2011. HapStar: Automated haplotype network layout and
655 visualisation. Mol. Ecol. Res. **11**:151-153.
656 Tennessen, J.A., Govindarajulu, R., Liston, A., and Ashman, T-L. 2013. Targeted sequence
657 capture provides insight into genome structure and genetics of male sterility in a
658 gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). Genes
659 Genomes Gen. **3**:1341-1351.
660 Tennessen, J.A., Govindarajulu, R., Liston A., and Ashman, T-L. 2014. Evolutionary origins and
661 dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage
662 maps. Genome Biol. Evol. **6**:3295–3313
663 Toews, D.P.L., and Brelsford, A. 2012. The biogeography of mitochondrial and nuclear
664 discordance in animals. Mol. Ecol. **21**:3907-3930.
665 Tsitrone, A., Kirkpatrick, M., and Levin, D. 2003. A model for chloroplast capture. Evolution
666 **57**:1776-1782.
667 Van Droogenbroeck, B., Kyndt, T.. Romeijn-Peeters, E., Van Thuyne, W., Goetghebeur, P.,
668 Romero-Motochi, J., and Gheysen, G. 2006. Evidence of natural hybridization and
669 introgression between *Vasconcellea species* (Caricaceae) from southern Ecuador revealed
670 by chloroplast, mitochondrial and nuclear DNA markers. Ann. Bot. **97**:793-805.
671 Vekemans, X., and Hardy, O.J. 2004. New insights from fine-scale spatial genetic structure
672 analyses in plant populations. Mol. Ecol. **13**:921-935.

Table 1

Genetic and sexual system variation across 15 populations of North American diploid strawberry

Taxon	Population Code	N	Mitotype	Mitotype Diversity	Chlorotype	Nuclear Membership		Allelic Richness	F_{IS}	Proportion Female	Mean Pollen Viability
			atp8-orf225 + *atp8*		*rpoC2*	Cluster 1	Cluster 2				
F. v. subsp. *bracteata*	OR-MRD	10, 9, 15, 37	C/B	0.20	2	0.004	0.996	2.58	0.15	0.3	0.75
F. v. subsp. *bracteata*	WA-DC	7, 7, 7, 20	B	0	2	0.004	0.996	2.00	0.13	0.1	0.81
F. v. subsp. *bracteata*	OR-FB	9, 10, 15, 9	C	0	2/1	0.009	0.991	2.13	0.46*	0.3	0.77
F. v. subsp. *bracteata*	CA-PDL	9, 10, 15, 20	B	0	2	0.011	0.989	1.94	0.08	0.05"	0.67
F. v. subsp. *bracteata*	CA-WMR	9, 10, 15, 22	B	0	2	0.011	0.989	2.07	0.09	0	0.87
F. v. subsp. *bracteata*	OR-ES	9, 10, 11, 38	G/C/E	0.64	1/2	0.009	0.991	2.20	0.29*	0.18	0.79
F. v. subsp. *bracteata*	ID-SFR	7, 9, 8, 9	E/C	0.46	1/2	0.004	0.996	1.60	0.27	0.1"	0.71
F. v. subsp. *bracteata*	CO-SCR	10, 12, 9, 11	C/G	0.41	1	0.034	0.966	1.46	0.66*	0.18"	0.73
F. v. subsp. *bracteata*	NM-LNF	9, 10, 18, 36	F	0	1	0.972	0.028	2.36	0.60*	0.46	0.80
F. v. subsp. *californica*	CA-INASD	2, 3, 4, 4	C	0	2	0.005	0.995	1.80	0.74*	0	0.90
F. v. subsp. *californica*	CA-SBNF	8, 10, 10, 18	C	0	2	0.005	0.995	1.20	N/A	0	0.85
F. mexicana	MX-LV	9, 10, 15, 27	C	0	1	0.004	0.996	1.58	0.62*	0	0.84
F. v. subsp. *americana*	IA-SSP	0, 4, 6, 6	na	na	1	0.979	0.021	1.53	0.07	0	0.95
F. v. subsp. *americana*	ONT-GRL	10, 10, 10, 10	D/H	0.46	1	0.789	0.211	2.12	0.76	0	0.94
F. v. subsp. *americana*	NH-FP	3, 3, 18, 18	H	0	1	0.996	0.004	1.39	0.26	0"	0.96

Note. Population code is given, see Table A1 for location details. N is the number of individuals used for the mitochondrial, chloroplast, and nuclear sequencing analyses, and for estimating proportion of females, respectively. Mitotype reflects the combination of sequence data for *atp8-orf225* and *atp8* (Fig. 2*b*, Table A2). Chlorotype is sequence data from *rpoC2* (Fig. 2*a*, Table A2). Order in polymorphic populations reflects majority/minority types. Nuclear membership in two clusters based on STRUCTURE analyses (Fig. 3). Allelic richness (number of alleles per locus) and inbreeding coefficients (F_{IS}). The proportion of females and pollen viability of hermaphrodites from each population.

"Determined from field-collected plants grown in the greenhouse. All other proportions estimated from plants observed in the field.

*Inbreeding coefficient is significant.

Na: not available.

Figure captions

Fig. 1. Schematic of the distributions of *Fragaria vesca* subspecies and *F. mexicana* in North America based on data from Staudt (1999) and Liston et al. (2014).

Fig.2. Minimum spanning networks of *F. vesca* haplotypes. (A) 351 bp of *rpoC2* cpDNA, (B) 776 bp of concatenated *atp8-orf225* and *atp8* mtDNA, (C) 351 bp of *atp8* mtDNA, (D) 425 bp of *atp8-orf225* mtDNA. Colors indicate taxa and shades within them populations. Population codes reflect state and population moniker as in Table A1. Letters correspond to mitotype (Table A3) and numbers to chlorotype. Black dots represent unknown ancestral haplotypes.

Fig.3. Geographic patterns of diversity in the cytoplasmic genes in North American diploid *F. vesca*. Large symbols represent deeply sampled populations, while small ones represent single individuals (see Table A1). The outer circle represents mitotype, with pattern representing the concatenated of *atp8* and *atp8-orf225* genes (Fig. 2B). Dashed lines indicate that mitotype is unknown. The inner circle shows chlorotype based on *rpoC2* (see Table A2). Squares represent chlorotypes reported in Njuguna et al. (2013).

Fig.4. STRUCTURE analysis of nuclear microsatellite loci showing population and individual membership to $K = 2$ genetic groups. Yellow represents cluster 1 (*F. v.* subsp. *americana* type). Blue represents cluster 2 (*F. v.* subsp. *bracteata* type). Each population is identified by the population code name (Table 1) followed by the taxon in parentheses (A= *F. v.* subsp. *americana*; B= *F. v.* subsp. *bracteata*; C= *F. v.* subsp. *californica*; M= *F. mexicana*). Cytoplasmic membership for each population is illustrated above. The outer circle represents

mitotype, with pattern representing the concatenated of *atp8* and *atp8-orf225* genes (Fig. 2B). The inner circle shows chlorotype based on *rpoC2* (see Table A2). Dashed lines indicate that mitotype is unknown.

Appendices

Table A1

Collection information for sampled populations/individuals of diploid North American strawberry

Taxon	Population/Accession Name	Population Code	State/Province, Country	Latitude (N)	Longitude (W)	N	Collection Type	DNA
F. v. subsp. *bracteata*	NCGR PI 616651*	.	British Columbia, Canada	49.150	-124.500	1	Leaf tissue	C
F. v. subsp. *bracteata*	Mary's Road	OR-MRD	Oregon, USA	44.496	-123.540	16	Whole plant	C, M, N
F. v. subsp. *bracteata*	Dungeness Campground	WA-DC	Washington, USA	48.139	-123.190	7	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 664439	FVB1	Oregon, USA	42.111	-122.709	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	Fishermen's Bend	OR-FB	Oregon, USA	44.756	-122.520	15	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 664438	FVB2	California, USA	41.352	-122.285	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	NCGR PI 551835*	.	Oregon, USA	44.500	-122.000	1	Leaf tissue	C
F. v. subsp. *bracteata*	NCGR PI 664413	FVB3	Oregon, USA	45.373	-121.823	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	Paradise Lake	CA-PDL	California, USA	39.859	-121.581	15	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 551807*	.	California, USA	39.782	-121.321	1	Leaf tissue	C
F. v. subsp. *bracteata*	NCGR PI 551749*	.	California, USA	34.504	-120.498	1	Leaf tissue	C
F. v. subsp. *bracteata*	White Meadow Road	CA-WMR	California, USA	38.774	-120.456	17	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 551551	FVB4	Oregon, USA	44.517	-120.371	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	Emigrant Springs	OR-ES	Oregon, USA	45.543	-118.460	11	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 551646*	.	Idaho, USA	48.538	-116.357	1	Leaf tissue	C
F. v. subsp. *bracteata*	NCGR PI 551644	FVB5	Idaho, USA	48.978	-116.157	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	South Fork Rd	ID-SFR	Idaho, USA	44.014	-115.512	9	Seed	C, M, N
F. v. subsp. *bracteata*	NCGR PI 551525	FVB6	Montana, USA	46.031	-114.171	1	Leaf tissue	C,M

F. v. subsp. *bracteata*	NCGR PI 651550*	.	Idaho, USA	42.163	-111.646	1	Leaf tissue	C
F. v. subsp. *bracteata*	NCGR PI 551523	FVB7	Wyoming, USA	43.634	-110.547	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	NCGR PI 551874	FVB8	Wyoming, USA	44.667	-110.500	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	NCGR PI 637952	FVB9	Arizona, USA	31.882	-109.284	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	Spring Creek Road	CO-SCR	Colorado, USA	38.762	-106.767	10	Seed	C, M, N
F. v. subsp. *bracteata*	Lincoln National Forest	NM-LNF	New Mexico, USA	32.968	-105.750	15	Whole plant	C, M, N
F. v. subsp. *bracteata*	NCGR PI 637947	FVB10	New Mexico, USA	35.703	-105.686	1	Leaf tissue	C,M
F. v. subsp. *bracteata*	NCGR PI 657860*	.	Colorado, USA	39.693	-105.501	1	Leaf tissue	C
F. v. subsp. *bracteata*	NCGR PI 551881*	.	South Dakota, USA	43.833	-103.500	1	Leaf tissue	C
F. v. subsp. *californica*	Inaja Picnic Area	CA-INASD	California, USA	33.023	-117.095	5	Whole plant	C, M, N
F. v. subsp. *californica*	San Bernadino National Forest	CA-SBNF	California, USA	34.163	-116.911	10	Whole plant	C, M, N
F. mexicana	Las Vigas	MX-LV	Veracruz, Mexico	19.539	-97.174	15	Seed	C, M, N
F. mexicana	Perote	MX-PE	Veracruz, Mexico	19.381	-97.090	1	Seed	C, M
F. v. subsp. *americana*	Springbrook State Park	IA-SSP	Iowa, USA	41.775	-94.465	6	Seed	C, M, N
F. v. subsp. *americana*	Guelph Radial Line	ONT-GRL	Ontario, Canada	43.549	-80.218	10	Leaf tissue	C, M, N
F. v. subsp. *americana*	Fish Pond Road	NH-FP	New Hampshire, USA	44.871	-71.504	15	Seed	C, M, N

Note. Taxonomic designations were based on Staudt (1999) except for the recognition of *F. mexicana*. Location is name of the collection site or accession number from the National Clonal Germplasm Repository. Population code is the abbreviated name used in the text. N is reported as the maximum number of plants used in sequencing analyses. Tissue collection type is reported as whole plant, leaf tissue (either dried or fresh) or seed (grown in greenhouse). DNA type shows which genomes (C= chloroplast, M = mitochondrial, N = nuclear) have been analyzed for each population/individual in the study. Asterisks indicate individuals with known sequences from Njuguna et al. 2013.

Table A2

Primer sequences for the mitochondrial genes and nuclear microsatellites screened to assess polymorphism

Genome	Gene/SSR Locus	Primer sequence	Expected product size	Annealing temperature (°C)	Source
Chloroplast	*rpoC2*	F: GGAATTCGAAATTCTCCCGTTT			
		R: AGGGATAATCTAGAGCTTCGAGTTG	700	52	Njuguna et al. 2013
Mitochondrial	*atp4*	F: CACGACTAGAAAGATGCTATTTGC			
		R: TAAGACCACCAAGCCCTCTC	526	54	Barr et al. 2007
Mitochondrial	*atp6*	F: CAATTTGCCATTCTCCCATT			
		R: TGATGGAGATTTGTAGCATCATTC	746	52	Barr et al. 2007
Mitochondrial	*atp8*	F: CTTCTGGTCATGCCTTTTCC			
		R: CTTGGCCGTGTGGAACAT	425	54	In this study
Mitochondrial	*atp9*	F: ATAGGTGCCGGAGCTGCTA			
		R: CGGAATACGGATGAGATCAAA	203	54	Darracq et al. 2012
Mitochondrial	*cob*	F: TGTTTGGTGTCTCGGAGTTG			
		R: TAACAAATGGTGCCTCCACA	713	54	Touzet and Delph 2009
Mitochondrial	*cox3*	F: CCAAGTCCATGGCCTATTTC			
		R: CAAATGGGAATAACCGAACC	730	52	Barr et al. 2007
Mitochondrial	*nad6*	F: TGCTTTGGTCTCTGGTTTGA			
		R: GTAGATCGTGACGGGGTCTG	589	54	Case and Willis 2008
Mitochondrial	*nad9*	F: TGGGAGACTTTACCTAAGAAATGG			
		R: TCCGTCGCTACGCTGTTC	540	54	Case and Willis 2008
Mitochondrial	*rps1*	F: TTTGAGTCGATCATTTCCAAGA			
		R: AGTAATGAAACCCCCGATGC	496	52	In this study
Mitochondrial	*atp8-orf225*	F: CATTTAATAATGCAGGGGTACG			
		R: CTGTGGTAGTCTTTCGTGG	425	48	In this study

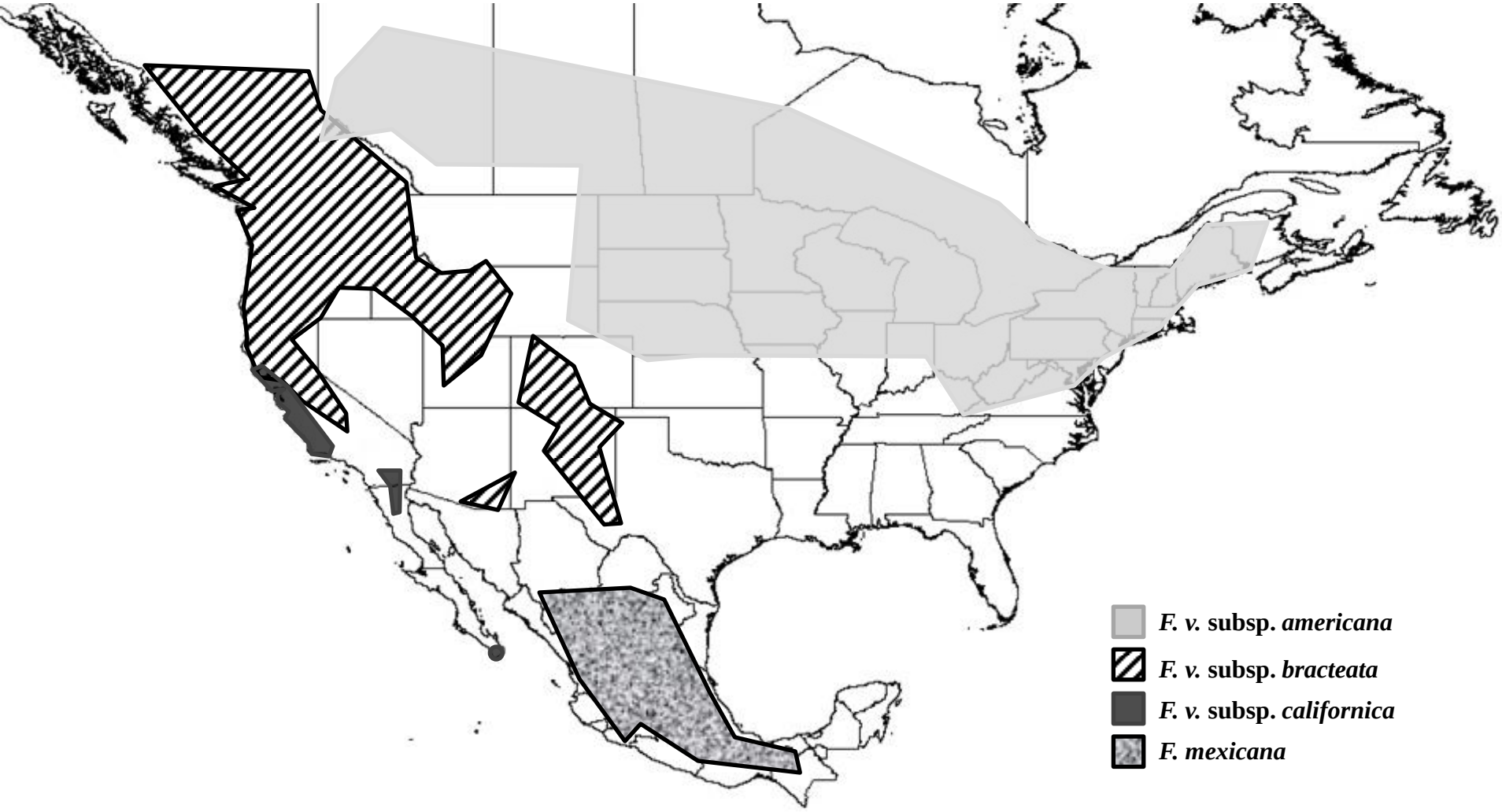
Nuclear	CX661264	F: GTTCTCAGATCCCTCTACCG			
		R: AATTTGCAGCCATCAAGTCC	483-487	59	Spigler et al. 2008
Nuclear	ARSFL007	F: GCGCGCATAAGGCAACAAAG			
		R: GCGAATGGCAATGACATCTTCTCT	227-238	59	Lewers et al. 2005
Nuclear	CX661603	F: ATCAACCACACCGCTACTCC			
		R: ATTTACGAAAATGCCATCGG	99-128	59	Spigler et al. 2008
Nuclear	UDF002	F: TATGGCCAGGATTGTTTGCT			
		R: TAGGAGGAGGCGTTGAAATG	132-148	59	Sargent et al. 2004
Nuclear	Fvi11	F: GCATCATCGTCATAATGAGTGC			
		R: GGCTTCATCTCTGCAATTCAA	320-358	59	Ashley et al. 2003

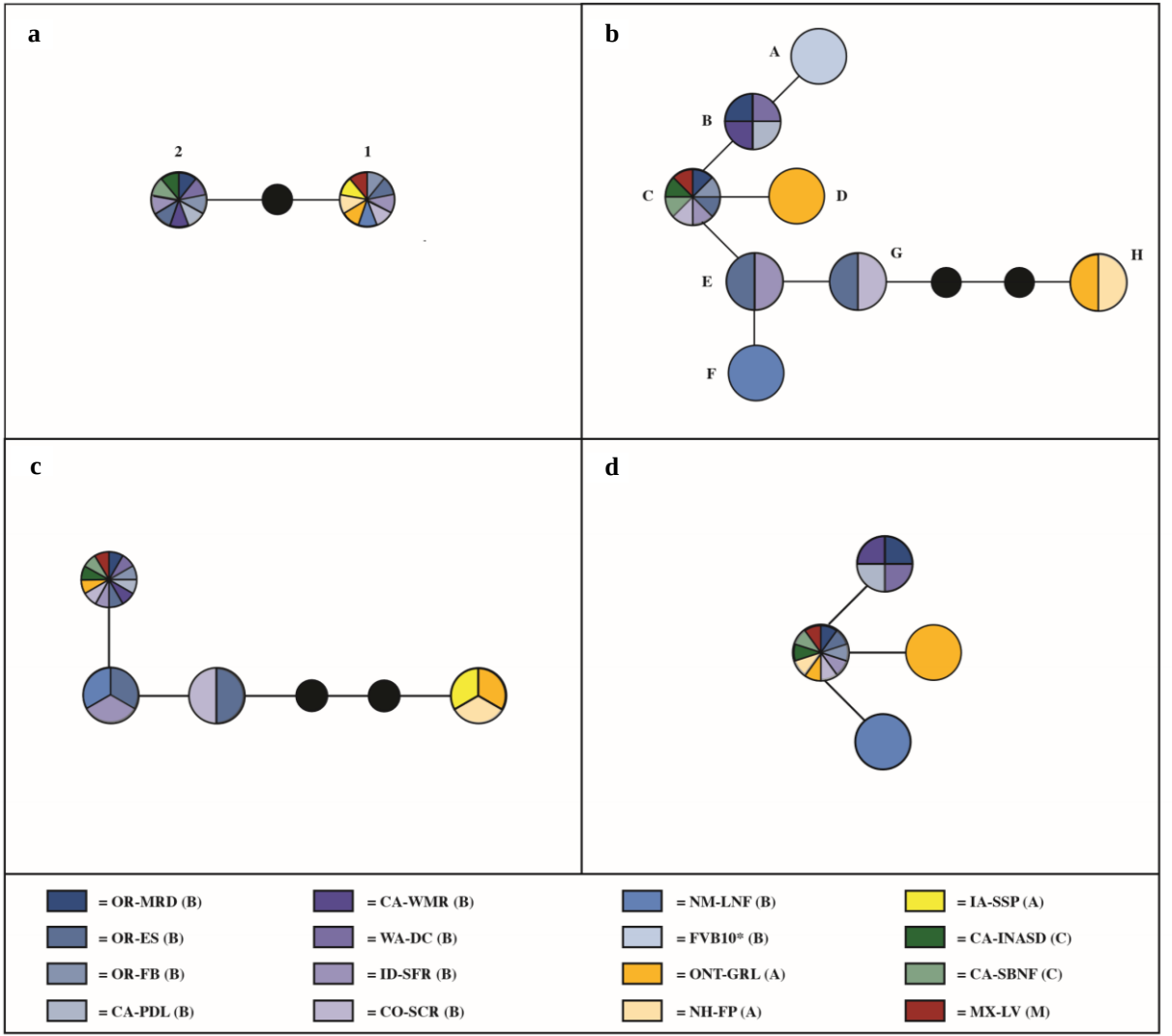
Table A3

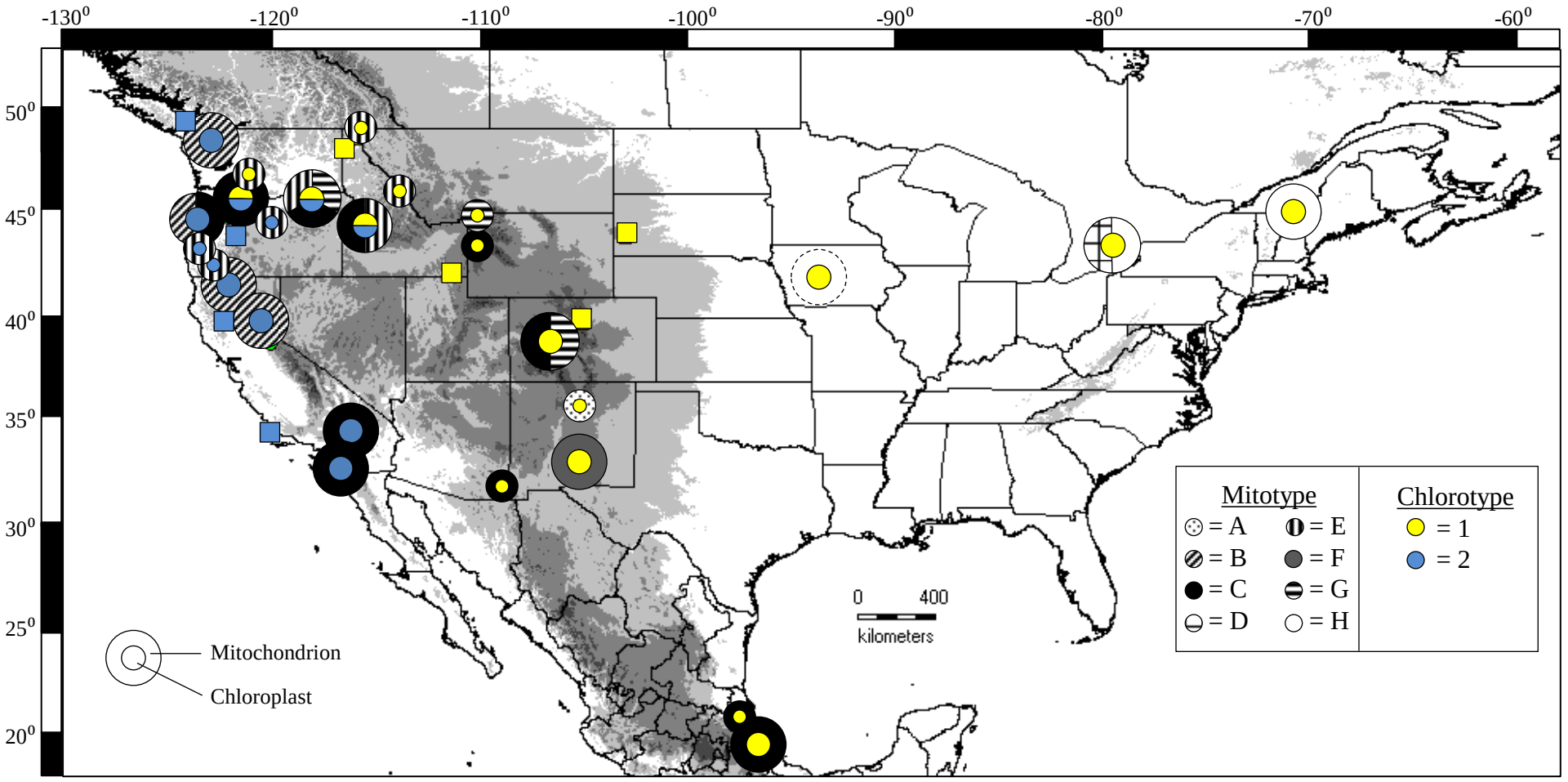
DNA polymorphism for *atp8-orf225* and *atp8* for four diploid *Fragaria* taxa.

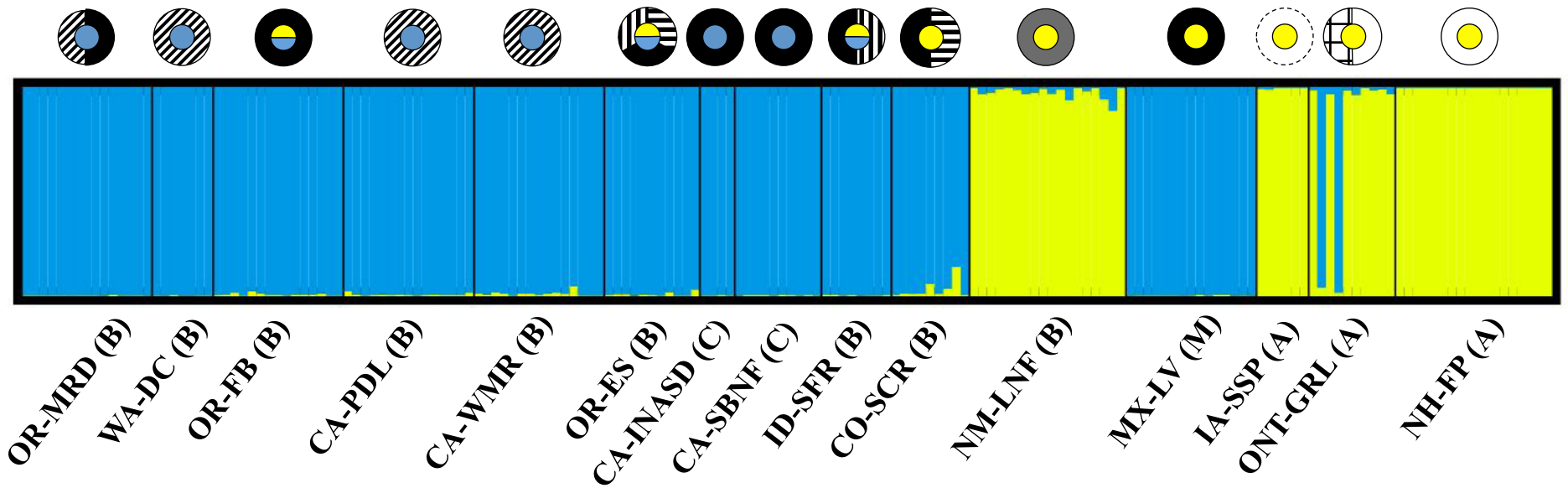
Gene	*atp8-orf225*			*atp8*				
Variable Locus	20	130	164	153	202	364	369	439
Mitotype								
A	G	C	G	G	A	G	G	C
B	T	A	G	G	A	G	G	C
C	T	C	G	G	A	G	G	C
D	T	C	T	G	A	G	G	C
E	T	C	G	G	A	G	T	C
F	G	C	G	G	A	G	T	C
G	T	C	G	G	A	T	T	C
H	T	C	G	T	C	T	T	T

Note. For each of the mitotypes (Table 1 and depicted in Fig.2) the variable sites in *atp8-orf225* and *atp8* (numbered from the first codon position) are given. Underline indicates position at which nonsynonymous substitutions occur. Mitotype A occurred in a single herbarium sample.


F. v. subsp. americana
F. v. subsp. bracteata
F. v. subsp. californica
F. mexicana


a
2
1
b
A
B
C
D
E
F
G
H
c
d
= OR-MRD (B)
= OR-ES (B)
= OR-FB (B)
= CA-PDL (B)
= CA-WMR (B)
= WA-DC (B)
= ID-SFR (B)
= CO-SCR (B)
= NM-LNF (B)
= FVB10* (B)
= ONT-GRL (A)
= NH-FP (A)
= IA-SSP (A)
= CA-INASD (C)
= CA-SBNF (C)
= MX-LV (M)


-130°
-120°
-110°
-100°
-90°
-80°
-70°
-60°
50°
45°
40°
35°
30°
25°
20°
Mitotype
= A
= B
= C
= D
= E
= F
= G
= H
Chlorotype
= 1
= 2
0
400
kilometers
Mitochondrion
Chloroplast


OR-MRD (B)
WA-DC (B)
OR-FB (B)
CA-PDL (B)
CA-WMR (B)
OR-ES (B)
CA-INASD (C)
CA-SBNF (C)
ID-SFR (B)
CO-SCR (B)
NM-LNF (B)
MX-LV (M)
IA-SSP (A)
ONT-GRL (A)
NH-FP (A)