

Genetic Variation and Distribution of Pacific Crabapple

Kanin J. Routson¹

University of Arizona, Arid Lands Resource Sciences, 1955 East Sixth Street, P.O. Box 210184, Tucson, AZ 85719

Gayle M. Volk and Christopher M. Richards

National Center for Genetic Resources Preservation, U.S. Department of Agriculture, 1111 S. Mason Street, Fort Collins, CO 80521

Steven E. Smith

School of Natural Resources and the Environment, University of Arizona, P.O. Box 210043, Tucson, AZ 85721

Gary Paul Nabhan

The Southwest Center and School of Geography and Development, University of Arizona, 1052 N. Highland Avenue, Tucson, AZ 85721

Victoria Wyllie de Echeverria

School of Environmental Studies, University of Victoria, P.O. Box 3060 STN CSC, Victoria, British Columbia V8W 3R4, Canada

ADDITIONAL INDEX WORDS. species distribution modeling, genetic diversity, *Malus fusca*, ex situ gene conservation

ABSTRACT. Pacific crabapple [*Malus fusca* (Raf.) C.K. Schneid.] is found in mesic coastal habitats in Pacific northwestern North America. It is one of four apple species native to North America. *M. fusca* is culturally important to First Nations of the region who value and use the fruit of this species as food, bark and leaves for medicine, and wood for making tools and in construction. However, little is known about either distribution or genetic diversity of this species. To correct this deficiency, we used habitat suitability modeling to map *M. fusca* habitat types with species occurrence records. The species apparently occupies at least two distinct climate regions: a colder, drier northern region and a warmer, wetter southern region. Total area of modeled habitat encompasses $\approx 356,780$ km² of low-lying areas along the Pacific coast. A total of 239 *M. fusca* individuals sampled from across its native range were genetically compared using six microsatellite markers to assess for possible geographic structuring of genotypes. The primers amplified 50 alleles. Significant isolation by distance was identified across the ≈ 2600 km (straight line) where samples were distributed. These results may help establish priorities for in situ and ex situ *M. fusca* conservation.

Wild apple species (*Malus* Mill.) are native throughout temperate climates of Asia, Europe, and North America (Brown, 2012; Luby, 2003). They offer promising sources of genetic diversity for apple breeding (Brown, 2012) and also provide a wildlife habitat and serve as a direct food source for humans. Four *Malus* species are native to North America. Three species occur in eastern and midwestern United States and eastern Canada: *M. angustifolia* (Aiton) Michx. is native from southern New Jersey to Florida, *M. coronaria* (L.) Mill. from Ontario to South Carolina, and *M. ioensis* Alph. Wood Britton ranges from

Minnesota to Texas. These have been determined to be closely related based on isozymes (Dickson et al., 1991). The species native to the Pacific northwestern North America, *M. fusca*, is the sole geographic, morphological (Van Eseltine, 1933), chemical (Williams, 1982), and genetic outlier among the North American taxa.

Pacific crabapple is a small, deciduous, often multitrunked tree or thicket-forming shrub that occurs naturally in mesic environments along river bottoms, meadows, and muskeg fringes at low to mid-elevations along the Pacific coast of North America from northern California to the Kenai Peninsula in Alaska (Viereck and Little, 1986). Yellow to red fruit are oblong and 1 to 2 cm in length (Fig. 1). It groups genetically with the species native to central Asia and China rather than the other North American taxa, according to amplified fragment length polymorphism data (Qian et al., 2006) and nuclear ribosomal and chloroplast DNA (Robinson et al., 2001). *M. fusca* is considered to be in section *Kansuenses* Rehd. along with *M. kansuensis* (Batalin) C.K. Schneid., *M. toringoides* (Rehder) Hughes, and *M. transitoria* (Batalin) C.K. Schneid. (Robinson et al., 2001). Because of genetic grouping with central Asia, the species is hypothesized to be a recent migrant across the Bering Strait (Williams, 1982).

Received for publication 29 May 2012. Accepted for publication 12 July 2012. We sincerely thank all who have contributed to the success of this research. We specifically acknowledge our British Columbia collaborators Nancy Turner, Leslie Main Johnson, and Ken Downs for their insights samples from the region. Thanks to Phillip Jenkins and Sarah Hunkins at the University of Arizona Herbarium (ARIZ) for assistance with herbarium records and Steffi Ickert-Bond at the University of Alaska Museum of the North Herbarium (ALA) and staff at the University of British Columbia Herbarium (UBC) and University of Alberta Herbarium (ALTA) for specimens. Thanks to Sarah Hayes, Joseph Postman, and all who helped with sample collection. Thanks to Adam Henk for technical assistance in the laboratory and Ned Garvey and Karen Williams for securing Plant Germplasm Collection funds. Finally, thanks to the Kellogg Program of the University of Arizona Southwest Center for funding.

¹Corresponding author. E-mail: kjr53@e-mail.arizona.edu.

Materials and Methods

SAMPLE COLLECTIONS. Two hundred thirty-nine *M. fusca* individuals were sampled across the species' native range from southeastern Alaska to northern California. Samples were obtained from field collections (138 individuals from California, Oregon, Washington, and British Columbia), herbarium specimens (65 individuals from Alaska and British Columbia), and U.S. Department of Agriculture (USDA) germplasm accessions (36 individuals from California, Oregon, and Washington). The latter were obtained from the USDA Agricultural Research Service Plant Genetic Resources Unit (PGRU) collection in Geneva, NY. The University of Alaska Museum of the North Herbarium and University of British Columbia Herbarium provided herbarium specimens.

Field collections took place Oct. 2010. Collections were made in publicly accessible parks, natural areas, and roadsides in moist, coastal habitats. Fresh leaf tissue was sampled (generally five leaves per tree) for genetic analysis. The fresh leaves were stored in plastic bags and kept cool until tissue samples were sent to the laboratory. There they were loaded into DNA extraction plates, after which they were frozen at -80°C for long-term storage. General physical site characteristics, habitat type, associated vegetation, and locality descriptions were documented and photographs taken at each site. Latitude/longitude coordinates and elevation were recorded for individual trees during field collections using a handheld GPS (eTrex-Vista; Garmin, Olathe, KS). Herbarium vouchers were collected at six sites and sent to the University of Arizona and University of Washington herbariums. Locations for herbarium specimens and USDA accessions were verified by comparing collection notes with Google Earth (Version 6.2; Google, Mountain View, CA) and biogeoreferencing web application GEOLocate Web (Rios and Bart, 2005). Points were selected based on habitat characteristics and descriptions of the localities. Specimens with less than a 5000-m uncertainty in GEOLocate were included in the species distribution modeling.

SPECIES DISTRIBUTION MODELING. A SDM for *M. fusca* was created using MaxEnt software [Version 3.3.3 (Phillips et al., 2006)] and WorldClim 1.4 spatially interpolated climate grids (Hijmans et al., 2005) and *M. fusca* collection localities ($n = 205$ with less than 5000-m uncertainty) and *M. fusca* occurrence records ($n = 152$ with less than 5000-m uncertainty) from the Global Biodiversity Information Facility (2001). We used 1950–2000 data at a 30-arc second resolution ($\approx 0.56\text{ km}^2$ at lat. 49°N) for elevation and 19 bioclimatic variables based on monthly precipitation and temperature. American Standard Code for Information Interchange (ASCII) files were generated in DIVA-GIS [Version 7.5.0 (Hijmans et al., 2001)] from WorldClim data for the Pacific northwestern North America region (long. 121.0°W to 151.0°W , lat. 40.0°N to 62.0°N).

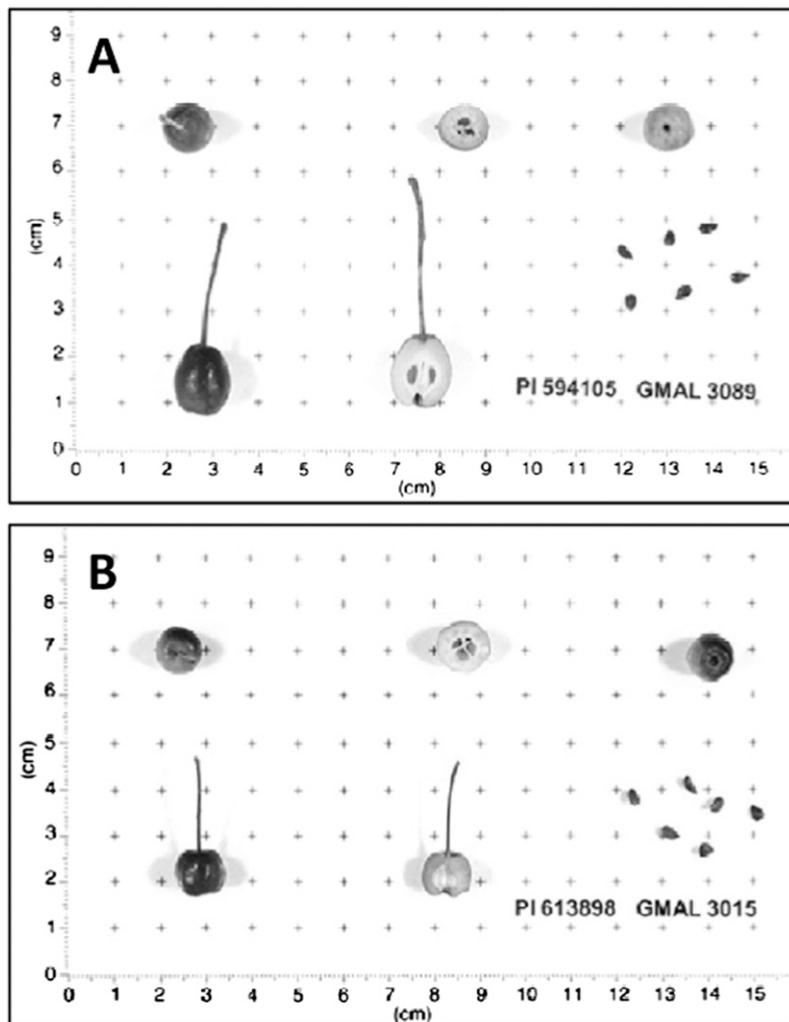


Fig. 1. Images of *Malus fusca* fruit collected from trees in (A) Washington and (B) California.

Malus fusca can serve as a rootstock for domesticated apple trees in waterlogged sites (R. Duncan, personal communication) and remains a culturally important species for First Nations of the Pacific northwestern North America (Downs, 2006; McDonald, 2005; Turner and Turner, 2008). All parts of the tree have been used including the fruit as food, the bark and leaves for medicines, and the dense wood for implements and in construction (Turner and Bell, 1971a, 1971b). It is well documented that Haida, Tsimshian, Tlingit, and Wakashan peoples in British Columbia and southeastern Alaska have tended small orchards of *M. fusca* that were often owned and managed in ways that were passed down between generations (Deur and Turner, 2005; Downs, 2006; McDonald, 2005; Turner and Peacock, 2005; Turner and Turner, 2008). Diploid *M. fusca* hybridizes with *M. ×domestica* Borkh. (Hartman, 1929) and has been a source of fire blight (*Erwinia amylovora* Burrill) resistance in apple breeding programs that use transgenic early flowering methods to reduce the length of the generation cycles (Flachowsky et al., 2011).

We describe genetic diversity and distribution of *M. fusca* through species distribution modeling (SDM) of known occurrence records and genetic fingerprinting of 239 *M. fusca* individuals sampled from northern California to southeastern Alaska using six microsatellite markers.

Multiple sample occurrences in grid cells ($n = 149$) were deleted using ENMTools (Warren et al., 2010) resulting in a total of 208 occurrence locations to build the model. A target group background occurrence file (Elith et al., 2011; Phillips et al., 2009) was constructed using 10,000 randomly selected occurrence records for flowering plants from the Global Biodiversity Information Facility (2001). Multiple occurrences within a grid were removed in ENMTools. We ran MaxEnt using default settings. A model was selected that 1) minimized Akaike information criterion scores [AICc (Warren and Seifert, 2011)] calculated in ENMTools over 10 replicate runs; 2) used environmental variables whose correlation was less than 0.60; 3) showed relatively high area under curve (AUC) scores; and 4) appeared consistent with known species occurrences. AUC scores relate to the probability of the model correctly scoring random presence and absence sites (Fielding and Bell, 1997; Phillips et al., 2009).

Malus fusca occurs across a wide latitudinal range and may occupy fundamentally different environments across this range (Nakazato et al., 2010). We assessed environmental variability across the *M. fusca* range by conducting a hierarchical cluster analysis using PROC CLUSTER (Ward's minimum-variance method) in SAS (Version 9.3; SAS Institute, Cary, NC). Observations were based on individual occurrences and clustering was performed on a subset of environmental variables selected to maximize variation and standardized to mean = 0 and SD = 1. We selected environmental variables by first performing a hierarchical of clustering of the variables themselves using PROC VARCLUS in SAS to identify correlation among variables and identify variables from non-overlapping clusters that maximized variation. Intrataxon clusters were identified by examining the cubic clustering criterion and the pseudo F and t2 statistics where $P \leq 0.01$ (Cooper and Milligan, 1988). Two percent of the observations with the lowest estimated probability density were omitted from clustering. In cases in which more than one cluster was identified, the omitted observations were grouped into the nearest cluster. Separate SDMs were developed for significantly distinct climate clusters.

Suitable habitat was quantified using Arcmap (ArcGIS 10; Esri, Redlands, CA). Probability of presence/background ASCII files produced by MaxEnt were converted into raster files in Arcmap. Map projections were converted from WGS1984 into Albers equal-area conic projection to calculate area of suitable habitat. The percentage of suitable habitat sampled in the genetic analyses was calculated by applying a 10-km buffer to each individual included in the genetic analyses, then merging buffered areas and clipping to exclude non-suitable habitat using the Buffer and Clip functions in Arcmap.

MICROSATELLITE ANALYSIS. The genetic analysis of the 239 samples was performed following procedures described in Volk et al. (2005). The six unlinked microsatellite markers used in this study can only represent 35% of the 17 linkage groups in the *Malus* genome. However, they have been used to successfully differentiate between individuals and identify structure in other *Malus* sp. (Richards et al., 2009; Volk et al., 2008). We extracted genomic DNA from frozen leaf tissue (field collections) and dried leaf tissue (herbarium records) using DNeasy 96 plant kits (Qiagen, Valencia, CA). DNA was extracted from the frozen leaves using 50 mg tissue frozen in liquid nitrogen during initial tissue grinding; 10 to 12 mg tissue was used for the dry samples and was extracted using the same protocol but without liquid nitrogen and a lower temperature of reagent

AP1. Unlinked primers (GD12, GD15, GD96, GD142, GD147, and GD162) described in Hemmat et al. (2003) and Hokanson et al. (1998) were used to amplify microsatellite loci. Infrared fluorescent dye IRD700 or IRD800 (MWG-Biotech, High Point, NC) labeled forward primers. Reverse primers were unlabeled (IDT, Coralville, IA).

Polymerase chain reactions (PCRs) were carried out in 15- μ L reactions. Each reaction contained: 0.25 μ L GoTaq® Flexi Taq Polymerase (Promega, Madison, WI) (5 U/ μ L), 3 μ L Promega 5 \times Colorless GoTaq® Flexi Buffer (10 mM Tris-HCl, 50 mM KCl, and 0.5% Triton X-100), 1.5 μ L of 0.25 mM MgCl₂, 1.5 μ L of 0.25 mM dNTPs, and 0.25 μ L forward and reverse primers. We added 5 μ L of undiluted genomic DNA product obtained from the Qiagen DNeasy 96 plant kits to each reaction. Addition of sterile distilled H₂O brought reaction volumes to 15 μ L. Primer pairs GD96, GD15, GD142, GD147, and GD162 were multiplexed. GD12 was run independently.

PCR was run on a PTC 200 Thermocycler (MJ Research, Reno, NV) using touchdown PCR. The annealing temperature was reduced 1 °C at each cycle, beginning at 63 °C and ending at 54 °C, followed by annealing at 55 °C for 18 cycles, and ending with a 72 °C extension for 2 min. PCR products were diluted 1:1 with a formamide bromophenol blue loading buffer and denatured at 95 °C for 5 min. The denatured products were run on gels (6.5% KB Plus acrylamide; LI-COR, Lincoln, NE) in 1 \times TBE buffer (89 mM Tris, 89 mM boric acid, and 20 mM EDTA) for 1 h, 45 min at 1500 V, 40 W, 40 mA, and 45 °C in a LI-COR 4200 DNA Sequencer. Digital images of the gels imported to LI-COR Saga Generation 2 software were visually analyzed in Saga. Overloaded gels were diluted 1:10 with additional loading buffer and rerun.

GENETIC DATA ANALYSIS. Allele frequencies, observed heterozygosity (H_o) and expected heterozygosity (H_e), allelic richness, and Wright's F-statistics were calculated in GDA (Lewis and Zaykin, 2002), FSTAT (Goudet, 1995), and Genodive (Meirmans and Van Tienderen, 2004) programs. Neighbor joining was computed in DARwin [Version 5.0.158 (Perrier and Jacquemoud-Collet, 2006)] from dissimilarity among genotypes (Perrier et al., 2003). Both dissimilarity calculations and neighbor joining were bootstrapped over 10,000 replications. Isolation by distance was determined by a Mantel (1967) test using Rousset's linearized F-statistic (F_{ST}), $\frac{F_{ST}}{1-F_{ST}}$ (Rousset, 1997), over 5000 permutations in Genodive.

We used STRUCTURE software (Pritchard et al., 2000) to identify possible evidence of population structure using posterior probabilities of possible populations. The number of populations was estimated from the natural log of the probability of allele frequencies over the posterior probability of one to 12 possible populations (Evanno et al., 2005). An ancestry model of admixture and correlated allele frequencies between populations enabled fractional assignment of individual genotypes to multiple populations by probability of membership. We ran a Markov chain Monte Carlo method in STRUCTURE using an iteration burn-in period of 500,000 runs followed by 100,000 iterations per chain. We ran each Markov chain 100 times for one to 12 possible populations. STRUCTURE output was captured and analyzed by STRUCTURE HARVESTER (Earl and vonHoldt, 2011). CLUMPP software produced population membership coefficients based on matrices of multiple STRUCTURE runs to average multiple population assignment runs for individuals belonging to more than one cluster (Jakobsson and Rosenberg, 2007). Individual genotypes were

assigned to the population cluster in which they had the highest membership.

Results and Discussion

SPECIES DISTRIBUTION MODELING. Cluster analysis of six bioclimatic variables on 157 georeferenced individuals identified two significant climatic clusters (Fig. 2), revealing a colder and drier “northern” cluster with 47 individuals (mean annual temperature = 4.2 °C, annual precipitation = 1702 mm) and a warmer and wetter “southern” cluster with 110 individuals (mean annual temperature = 9.1 °C, annual precipitation = 2267 mm). Bioclimatic variables of temperature seasonality, mean temperature of wettest quarter, mean temperature of warmest quarter, annual precipitation, precipitation of driest month, and precipitation of coldest quarter were used in the cluster analyses. These variables were selected from the PROC VARCLUS hierarchical of clustering on the variables themselves to identify variables from non-overlapping clusters that maximized variation (see “Materials and Methods”). The two climatic clusters are not perfectly segregated in space. Two species occurrences on the north end of Haida Gwaii, British Columbia, Canada, were characterized as being climatically grouped with the “southern” cluster, although they are generally well within the “northern” climatic region. Ocean currents or some “island” phenomenon associated with Haida Gwaii could be responsible for the mixing of clusters.

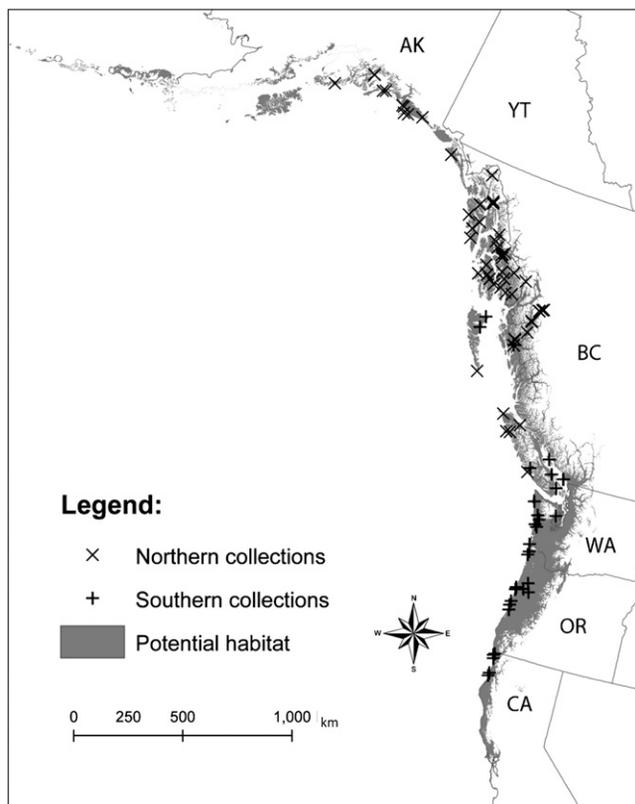


Fig. 2. A habitat suitability model derived from presence only data in MaxEnt indicates predicted suitable habitat for *Malus fusca* in the shaded areas. Northern (“x” signs) and Southern (“+” signs) denote two significantly different habitat types derived from clustering of six WorldClim bioclimatic variables for *M. fusca* presence of data based on collection localities.

Potential *M. fusca* habitat was mapped using two SDMs for the northern and southern climatic clusters (Fig. 2). The SDM for the northern cluster for *M. fusca* was developed using eight WorldClim bioclimatic variables. Bioclimatic variables of altitude, mean diurnal temperature range, minimum temperature of coldest month, mean temperature of driest quarter, mean temperature of warmest quarter, annual precipitation, precipitation of driest month, and precipitation of coldest quarter were included as model elements. The SDM for the southern climatic cluster used seven bioclimatic variables including: altitude, annual mean temperature, mean diurnal temperature range, mean temperature of wettest quarter, annual precipitation, precipitation of warmest quarter, and precipitation of coldest quarter as model elements. The models were selected based on low AICc scores [6134 (northern), 5708 (southern)] calculated over 10 replicate runs, less than 0.60 correlation among environmental variables, high AUC scores (AUC = 0.985 in the northern cluster, AUC = 0.986 in the southern cluster), and consistency with known species occurrences. When the two SDMs were merged in ArcMap, there is 356,780 km² of potential habitat for *M. fusca* (Fig. 2). Using a buffer area of 10 km, samples representing 15,306 km² of potential habitat were used in the genetic analyses, equating to 4.29% of the potential habitat.

GENETIC DATA ANALYSIS. The six primers amplified a total of 50 alleles (Table 1). Polymorphic alleles observed in the 239 individuals ranged from a low of two for GD15 to a maximum of 20 for GD147. Observed heterozygosity (0.343) was found to be lower than expected heterozygosity (0.428) suggesting slight inbreeding may be occurring (Table 1). In another apple species, *M. sieversii* (Ledeb.) M. Roem., Richards et al. (2009) report much higher values of observed and expected heterozygosities (He = 0.749; Ho = 0.693). This may indicate higher total genetic diversity in *M. sieversii* compared with *M. fusca*.

The six primers were sufficient to differentiate among all individuals except nine pairs of genotypes. Seven of these were from adjacent trees and two from herbarium records. Two duplicate genotypes were found in the University of Alaska herbarium specimens, one from the vicinity of Katalia, AK, and from the Yakutat Foreland, AK, which are ≈300 km apart. Another herbarium specimen’s genotype from the Alexander Archipelago, AK, matched a genotype from the Skeena River, near Kitwanga, British Columbia, Canada, also more than 300 km distant.

An additional seven markers, described by Liebhard et al. (2002), were run on the nine pairs of duplicates and 27 individuals randomly selected from the northern (15 individuals) and southern clusters (12 individuals) based on percentage of individuals sampled from unique grid cells in each habitat to confirm the results of the six markers (Table 1). These additional markers resolved the geographically isolated duplicates into unique genotypes. Two of the duplicates from adjacent trees were not resolved and are presumed to have been collected from single individuals (the shrub nature of *M. fusca*, possible suckering, and often densely vegetated habitat make differentiating between individual trees difficult in some cases).

When individuals collected within 1 km of each other were assumed to be from the same population and grouped together, we found significant differences between populations using analysis of molecular variance (Excoffier et al., 1992; Michalakis and Excoffier, 1996) with an $F_{ST} = 0.074$, $P = 0.001$ (range, 0.05 to 0.099). However, there is a much higher within-population

variability (F-statistic $F_{IS} = 0.135$, $P = 0.001$). Overall, these values are similar to those reported for other species (Table 2), except that fewer polymorphic alleles were identified in *M. fusca* than either *M. sieversii* (Richards et al., 2009) or *M. orientalis* Uglitzk. (Volk et al., 2008). This may indicate a lower total genetic variation in *M. fusca* compared with other wild *Malus* species.

A neighbor-joining dendrogram (Fig. 3) of the 47 individuals from the northern climatic region and 47 individuals randomly selected individuals from the southern climatic cluster show no distinct separation of clusters by region. This suggests high admixture among *M. fusca* and a general lack of genetic structure.

We found low but significant isolation by distance (IBD) for $\frac{F_{ST}}{1-F_{ST}}$ by kilometers between populations (defined as all individuals within 1 km of each other). Mantel's r was calculated at 0.395 ($P = 0.007$, error sum of squares = 0.483, $R^2 = 0.156$). Isolation by distance reported in other species of woody Rosaceae range from marginal isolation by distance ($R^2 = 0.17$, $P = 0.091$) across ≈ 10 km in *Prunus mahaleb* L. populations in southeastern Spain (Jordano and Godoy, 2000) to high isolation by distance ($R^2 = 0.67$, $P = 0.01$) across ≈ 2.5 km in island populations of wild flowering cherry [*Prunus lannesiana* (Carriere) E.H. Wilson] from the Izu Islands in Japan (Kato et al., 2011). Grouping of samples in an IBD plot indicates geographic barriers to gene flow (Guillot, 2009). The IBD plot for *M. fusca* (Fig. 4) does not show distinct groupings between samples, suggesting a single, continuous population under IBD. Although significant, these values of IBD seem very low compared with the magnitude of the distance surveyed (2600 km).

This may be a product of the high admixture observed in the species. Another possibility is that habitat-related selection or a previous founder event drove the species toward uniformity. It is also possible that not enough of the genome was sampled or too few individuals were sampled to accurately capture IBD structure. *M. fusca* is an animal-pollinated and dispersed species with the potential for effective long-distance dispersal. This, in combination with a relatively continuous current distribution of suitable habitat (Fig. 2), lends support to the genetic findings of a single continuous population with low IBD across the species range.

The 239 individuals were grouped into "population" clusters using posterior probabilities of allele frequencies in STRUCTURE software (Pritchard et al., 2000). Structure results identified four significant clusters. The composition of individuals within clusters is highly admixed both genetically and geographically. Each individual was assigned to multiple clusters (the highest percentage of any individual belonging to a single cluster was 41% with an average highest contribution of 33% across all samples). Clusters showed no geographic separation when individuals were plotted on Google Earth (Version 6.2; Google).

Pleistocene ice sheets in the Pacific northwestern North America [30,000 to 10,000 C years BP (uncalibrated radiocarbon date); Clague and James, 2002] resulted in strong geographic structuring of many species now found in the region (reviewed in Shafer et al., 2010). Although this may have influenced the genetic structure of *M. fusca*, either through the extirpation of some isolated populations or through recolonization from isolated refugia, the high admixture and lack of discernible

Table 1. Descriptive statistics by locus.^z

Locus	Individuals (no.)	Polymorphic alleles (no.)	He	Ho	LG
GD12	237/27	5/5	0.538/0.567	0.544/0.481	3
GD15	237/27	2/2	0.159/0.307	0.127/0.296	Unknown
GD96	237/27	7/6	0.311/0.361	0.309/0.370	17
GD142	236/27	10/2	0.122/0.037	0.122/0.037	9
GD147	236/27	20/13	0.859/0.866	0.410/0.556	13
GD162	239/27	6/5	0.580/0.633	0.544/0.444	4
CH01d09	27	16	0.936	0.741	12
CH02b12	27	9	0.846	0.852	5
CH02d08	26	5	0.545	0.577	11
CH05e03	21	24	0.962	0.905	2
NH009b	27	4	0.640	0.370	13
NH015a	22	15	0.933	0.636	Unknown
NZ28f4	27	6	0.676	0.593	12
Mean	237/26.1	10.2/8.6	0.428/0.639	0.343/0.528	—
Total	239/27	50/112	—	—	—

^zSix markers were run across 239 individuals, and 13 markers were run on 27 randomly selected individuals of *Malus fusca*. Reported values include number of individuals, number of polymorphic alleles, expected heterozygosity (He), observed heterozygosity (Ho), and mapped linkage group (LG) for each marker (Celton et al., 2009).

Table 2. Diversity comparisons with other *Malus* sp. Geographic range (maximum distance), number of markers in the analyses, number of amplified alleles, within population (pop.) variance (F_{IS}), and among population variance (F_{ST}) are presented for *M. fusca*, *M. sieversii* (Richards et al., 2009), and *M. orientalis* (Volk et al., 2008).

Species	Geographic range (km)	Markers (no.)	Alleles (no.)	F_{IS}	F_{ST}	Reference
<i>M. fusca</i>	≈ 2600	6	50	0.135	0.074	Present study
<i>M. sieversii</i>	≈ 1100	7	103	0.190	0.050	Richards et al., 2009
<i>M. orientalis</i>	≈ 760	7	115	0.208	0.046	Volk et al., 2008

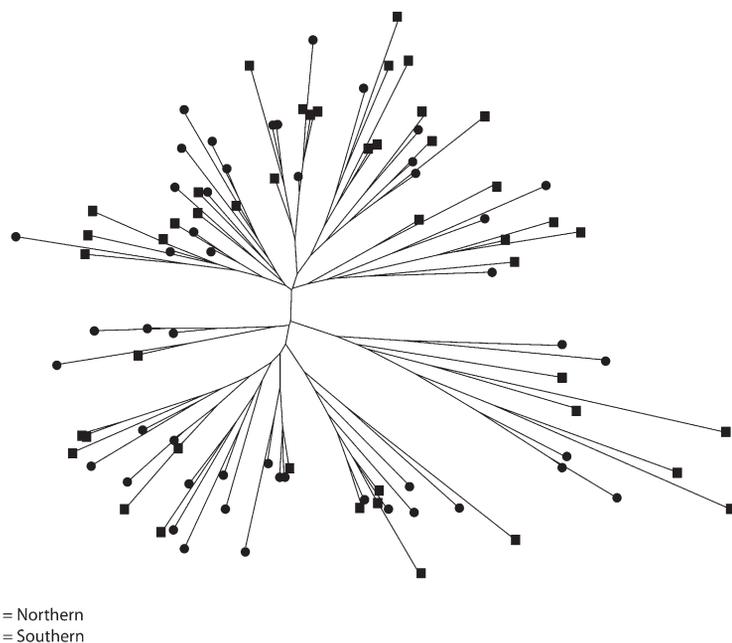


Fig. 3. Interindividual distances among *Malus fusca* genotypes. Individuals are labeled by northern (squares) and southern (circles) climatic region where each individual was collected. The lack of distinct grouping between climatic region by cluster signifies high admixture in *M. fusca*.

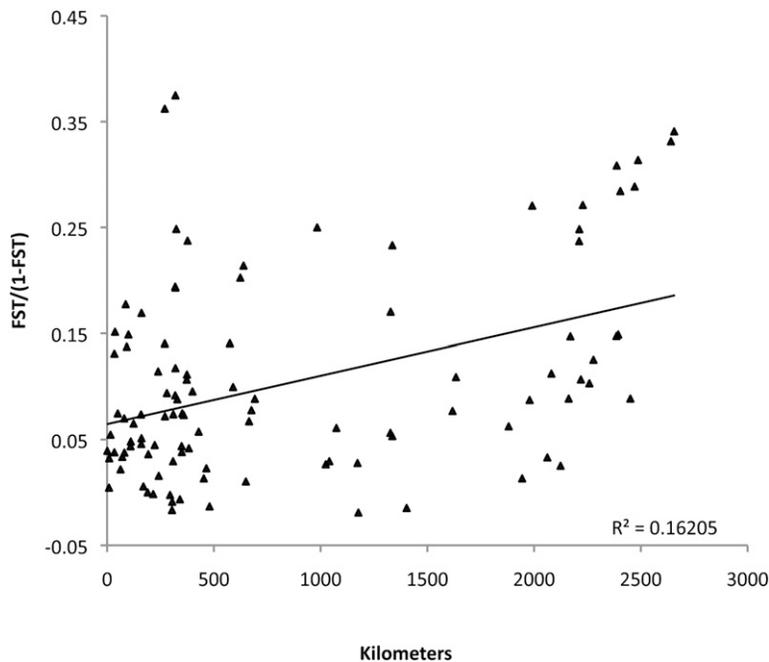


Fig. 4. Isolation by distance regression. *Malus fusca* shows significant isolation by distance as demonstrated in this biplot displaying geographical distance between sampling sites and genetic distance using Rousset's distance measure (Rousset, 1997). Individuals collected 1 km or less apart were grouped into populations. F_{st} is a measure of genetic differentiation between groups of genotypes ($R^2 = 0.162$, $P = 0.0019$).

geographic patterning in STRUCTURE results suggest a lack of geographic structuring of *M. fusca* populations, lending support to a recent introduction of the species in the last $\approx 10,000$ years. Other mechanisms could also account for a lack of geographic structuring and high admixture

including effective dispersal mechanisms or movement of plant material by humans or animals.

The USDA PGRU conserves *M. fusca* germplasm ex situ as 40 living accessions obtained from California, Oregon, Washington, and one from Alaska. Thirty-four of the 50 alleles found in the field collections and herbarium specimens are present in the USDA accessions. An additional 11 alleles present in the USDA accessions were not identified in field collections or herbarium records. These additional alleles could be present in wild populations not sampled or could be hybrid introgressions. F_{ST} between USDA collections and the herbarium specimens and the field collections showed slight, significant differences between the sample sets ($F_{ST} = 0.006$, $P = 0.043$), which suggest that USDA PGRU *M. fusca* collections generally represent wild populations, at least in the regions sampled. However, the USDA collection contains only a single representative sample from the northern climatic cluster, and although large differences in allele frequencies between the climatic clusters were not identified in the neutral markers used in this study, a comprehensive collection effort of the species may warrant additional collections from the northern climatic cluster.

Conclusions

We estimate suitable *M. fusca* habitat to be geographically limited to $\approx 356,780$ km². The habitats within this range consist of low-lying, moist areas along the northern Pacific coast. In the regions sampled (less than 5% of suitable habitat), *M. fusca* shows high admixture with little population differentiation, although it does show significant IBD across the ≈ 2600 km (straight line) sampled. High admixture in *M. fusca* could be a result of a continuous habitat distribution, effective pollen and seed dispersal mechanism, recent introduction and rapid dispersal of the species, trade or movement of *M. fusca* plant material by humans, or a combination of factors. This research can be brought to bear on ex situ and in situ conservation management of the species. USDA germplasm collection currently contains only a single representative sample of *M. fusca* from the northern regions of its range. Differences in the climate regions could justify additional collections in the northern region, although

the PGRU does capture much of the identified variability. Much of the *M. fusca* range covers regions not readily threatened by immediate coastal development or urban expansion. On the other hand, warmer temperatures and higher summer moisture deficits predicted under climate model simulations (Littell

et al., 2010; Mote and Salathé, 2010) could negatively affect *M. fusca* distribution in the future.

Literature Cited

- Brown, S.K. 2012. Apple, p. 329–367. In: Badenes, M.L. and D.H. Byrne (eds.). *Fruit breeding*. Springer, New York, NY.
- Celton, J.M., D. Tustin, D. Chagné, and S. Gardiner. 2009. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and *Pyrus* genomic sequences. *Tree Genet. Genomes* 5:93–107.
- Clague, J.J. and T.S. James. 2002. History and isostatic effects of the last ice sheet in southern British Columbia. *Quat. Sci. Rev.* 21:71–87.
- Cooper, M.C. and G.W. Milligan. 1988. The effect of measurement error on determining the number of clusters in cluster analysis, p. 319–328. In: Gaul, W. and M. Schader (eds.). *Data, expert knowledge, and decisions*. Springer-Verlag, London, UK.
- Deur, D. and N.J. Turner. 2005. Reconstructing indigenous resource management, reconstructing the history of an idea, p. 3–34. In: Deur, D. and N.J. Turner (eds.). *Keeping it living*. University of Washington Press, Seattle, WA.
- Dickson, E.E., S. Kresovich, and N.F. Weeden. 1991. Isozymes in North American *Malus* (Rosaceae): Hybridization and species differentiation. *Syst. Bot.* 16:363–375.
- Downs, K. 2006. The Tsimshian homeland: An ancient cultural landscape. Master's thesis, Athabasca Univ., Athabasca, Alberta, Canada.
- Earl, D.A. and B.M. vonHoldt. 2011. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet. Resources*. Version: V0.6.8, Oct. 2011.
- Elith, J., S.J. Phillips, T. Hastie, M. Dudik, Y.E. Chee, and C.J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17:43–57.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation summary. *Mol. Ecol.* 14:2611–2620.
- Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes—Application to human mitochondrial-DNA restriction data. *Genetics* 131:479–491.
- Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24:38–49.
- Flachowsky, H., P.-M. Le Roux, A. Peil, A. Patocchi, K. Richter, and M.-V. Hanke. 2011. Application of a high-speed breeding technology to apple (*Malus × domestica*) based on transgenic early flowering plants and marker-assisted selection. *New Phytol.* 192:364–377.
- Global Biodiversity Information Facility. 2001. Search species: *Malus fusca* (Raf.) C.K. Schneid. 5 Feb. 2012. <<http://data.gbif.org/species/3001080/>>.
- Goudet, J. 1995. FSTAT, a program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-statistics. *J. Hered.* 86:485–486.
- Guillot, G. 2009. On the inference of spatial structure from population genetics data. *Bioinformatics* 25:1796–1801.
- Hartman, H. 1929. Hybrids between *Pyrus malus* and *Pyrus fusca*. *J. Hered.* 20:379–390.
- Hemmat, M., S.K. Brown, and N.F. Weeden. 2003. Mapping and evaluation of *Malus × domestica* microsatellites in apple and pear. *J. Amer. Soc. Hort. Sci.* 128:515–520.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Intl. J. Climatol.* 25:1965–1978.
- Hijmans, R.J., L. Guarino, M. Cruz, and E. Rojas. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genet. Resour. Newsl.* 127:15–19.
- Hokanson, S.C., A.K. Szewc-McFadden, W.F. Lamboy, and J.R. McPerson. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus × domestica* Borkh. core collection. *Theor. Appl. Genet.* 97:671–683.
- Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
- Jordano, P. and J.A. Godoy. 2000. RAPD variation and population genetic structure in *Prunus mahaleb* (Rosaceae), an animal-dispersed tree. *Mol. Ecol.* 9:1293–1305.
- Kato, S., Y. Tsumura, H. Iwata, and Y. Mukai. 2011. Genetic structure of island populations of *Prunus lannesiana* var. *speciosa* revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. *J. Plant Res.* 124:11–23.
- Lewis, P.O. and D. Zaykin. 2002. GDA user's manual. 12 July 2012. <<http://www.eeb.uconn.edu/people/plewis/software.php>>.
- Liebhart, R., L. Gianfranceschi, B. Koller, C.D. Ryder, R. Tarchini, E. Weg, and C. Gessler. 2002. Development and characterization of 140 new microsatellites in apple (*Malus × domestica* Borkh.). *Mol. Breed.* 10:217–241.
- Littell, J.S., E.E. Oneil, D. McKenzie, J.A. Hicke, J.A. Lutz, R.A. Norheim, and M.M. Elsner. 2010. Forest ecosystems, disturbance, and climatic change in Washington State, USA. *Clim. Change* 102:1–2.
- Luby, J.J. 2003. 1. Taxonomic classification and brief history, p. 1–14. In: Ferree, D.C. and I.J. Warrington (eds.). *Apples: Botany, production and uses*. CABI, Cambridge, MA.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.
- McDonald, J.A. 2005. Cultivating in the northwest: Early accounts of Tsimshian horticulture, p. 240–273. In: Deur, D. and N.J. Turner (eds.). *Keeping it living*. University of Washington Press, Seattle, WA.
- Meirmans, P.G. and P.H. Van Tienderen. 2004. Genotype and genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4:792–794.
- Michalakis, Y. and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142:1061–1064.
- Mote, P.W. and J.E.P. Salathé. 2010. Future climate in the Pacific Northwest. *Clim. Change* 102:29–50.
- Nakazato, T., D.L. Warren, and L.C. Moyle. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *Amer. J. Bot.* 97:680–693.
- Perrier, X., A. Flori, and F. Bonnot. 2003. Data analysis methods, p. 43–76. In: Hamon, P., M. Seguin, X. Perrier, and J.C. Glaszmann (eds.). *Genetic diversity of cultivated tropical plants*. Enfield, Montpellier, France.
- Perrier, X. and J.P. Jacquemoud-Collet. 2006. DARwin software. 5 July 2012. <<http://darwin.cirad.fr/>>.
- Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190:231–259.
- Phillips, S.J., M. Dudik, J. Elith, C.H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* 19:181–197.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Qian, G.Z., L.F. Lui, and G.G. Tang. 2006. A new selection of *Malus* (Rosaceae) from China. *Ann. Bot. Fenn.* 43:68–73.
- Richards, C.M., G.M. Volk, A.A. Reilley, A.D. Henk, D. Lockwood, P.A. Reeves, and P.L. Forsline. 2009. Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated apple. *Tree Genet. Genomes* 5:339–347.
- Rios, N.E. and H.L. Bart, Jr. 2005. GEOLocate. Georeferencing software for natural history collections. 5 July 2012. <<http://www.museum.tulane.edu/geolocate/web/WebGeoref.aspx>>.

- Robinson, J.P., S.A. Harris, and B.E. Juniper. 2001. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst. Evol.* 226:35–58.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228.
- Shafer, A.B.A., C.I. Cullingham, D.W. Coltman, and S.D. Cote. 2010. Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Mol. Ecol.* 19:4589–4621.
- Turner, N.C. and M.A.M. Bell. 1971a. The ethnobotany of the Coast Salish Indians of Vancouver Island. *Econ. Bot.* 25:63–104.
- Turner, N.C. and M.A.M. Bell. 1971b. The ethnobotany of the Coast Salish Indians of Vancouver Island. Appendix I. *Econ. Bot.* 25:335–339.
- Turner, N.J. and S. Peacock. 2005. Solving the perennial paradox: Ethnobotanical evidence for plant resource management on the northwest coast, p. 101–150. In: Deur, D. and N.J. Turner (eds.). *Keeping it living*. University of Washington Press, Seattle, WA.
- Turner, N.J. and K.L. Turner. 2008. ‘Where our women used to get the food’: Cumulative effects and loss of ethnobotanical knowledge and practice; case study from coastal British Columbia. *Botany* 86:103–115.
- Van Eseltine, G.P. 1933. Notes on the species of apples. I. The American crabapples. *Agr. Expt. Sta. New York Tech. Bul. No.* 208.
- Viereck, L.A. and E.L. Little. 1986. Oregon crab apple (*Malus fusca*), p. 205–207. In: *Alaska trees and shrubs*. University of Alaska Press, Fairbanks, AK.
- Volk, G.M., C.M. Richards, A.A. Reilley, A.D. Henk, P.L. Forsline, and H.S. Aldwinckle. 2005. Ex situ conservation of vegetatively propagated species: Development of a seed-based core collection for *Malus sieversii*. *J. Amer. Soc. Hort. Sci.* 130:203–210.
- Volk, G.M., C.M. Richards, A.A. Reilley, A.D. Henk, P.A. Reeves, P.L. Forsline, and H.S. Aldwinckle. 2008. Genetic diversity of wild *Malus orientalis* from Turkey and southern Russia. *J. Amer. Soc. Hort. Sci.* 133:383–389.
- Warren, D.L., R.E. Glor, and M. Turelli. 2010. ENMTools: A toolbox for comparative studies of species distribution models. *Ecogeography* 33:607–611.
- Warren, D.L. and S.N. Seifert. 2011. Species distribution modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21:335–342.
- Williams, A.H. 1982. Chemical evidence from the flavonoids relevant to the classification of *Malus* species. *Bot. J. Linn. Soc.* 84:31–39.