RESEARCH ARTICLE

Patterns of Gene Flow between Crop and Wild Carrot, *Daucus carota* (Apiaceae) in the United States

Jennifer R. Mandel[1,2]*, **Adam J. Ramsey**[1], **Massimo Iorizzo**[3], **Philipp W. Simon**[4]

1 Department of Biological Sciences, The University of Memphis, Memphis, Tennessee, United States of America, 2 W. Harry Feinstone Center for Genomic Research, The University of Memphis, Memphis, Tennessee, United States of America, 3 Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University, Kannapolis, North Carolina, United States of America, 4 USDA-Agricultural Research Service, Vegetable Crops Unit, University of Wisconsin-Madison, Wisconsin, United States of America

* jmandel@memphis.edu

Abstract

Studies of gene flow between crops and their wild relatives have implications for both management practices for cultivation and understanding the risk of transgene escape. These types of studies may also yield insight into population dynamics and the evolutionary consequences of gene flow for wild relatives of crop species. Moreover, the comparison of genetic markers with different modes of inheritance, or transmission, such as those of the nuclear and chloroplast genomes, can inform the relative risk of transgene escape via pollen versus seed. Here we investigate patterns of gene flow between crop and wild carrot, *Daucus carota* (Apiaceae) in two regions of the United States. We employed 15 nuclear simple sequence repeat (SSR) markers and one polymorphic chloroplast marker. Further, we utilized both conventional population genetic metrics along with Shannon diversity indices as the latter have been proposed to be more sensitive to allele frequency changes and differentiation. We found that populations in both regions that were proximal to crop fields showed lower levels of differentiation to the crops than populations that were located farther away. We also found that Shannon measures were more sensitive to differences in both genetic diversity and differentiation in our study. Finally, we found indirect evidence of paternal transmission of chloroplast DNA and accompanying lower than expected levels of chloroplast genetic structure amongst populations as might be expected if chloroplast DNA genes flow through both seed and pollen. Our findings of substantial gene flow for both nuclear and chloroplast markers demonstrate the efficiency of both pollen and seed to transfer genetic information amongst populations of carrot.

Introduction

Cultivated plant species and their sexually-compatible wild relatives often overlap in terms of geographic proximity and phenology [1]. This overlap provides the opportunity for gene flow

between crops and their wild relatives [2,3]. Farmers and breeders are often concerned with the potential for wild allelic contamination *into* agricultural fields, which can hinder production efficiency. While, natural land managers may wish to minimize gene flow *out* of agricultural fields into natural, often weedy, populations of closely related species. Moreover, given a genetically modified (GM; a.k.a. genetically engineered) crop system, an additional consideration is the risk of escape of transgenes from an agricultural setting to the wild by gene flow through both hybridization, and/or introgression [4–7].

Gene flow between non-GM crops and their wild relatives has the potential to increase weediness of wild populations (see examples in relatives of canola, millet, radish rice, rye, sugar beet, and sorghum [8–11]). Moreover, increases in invasiveness and accompanying range expansions of a wild species as a result of introgression with crop relatives has been investigated in a number of species [12–17]. Given the growing knowledge that escaped transgenes, especially those conferring insect or pathogen resistance, could have detrimental effects on natural communities, there has been much interest in understanding patterns of gene flow between crops and their wild relatives [18–22]. Moreover, the recent announcement that genetically engineered crops are not harmful to human health combined with rising interest in genome editing technologies, could lead to an increase in the production of engineered crops in the near future and a greater focus on the potential for gene escape [23].

One particularly weedy crop wild relative is wild carrot, *Daucus carota* subsp. *carota* L., (a.k.a. Queen Anne's Lace). Wild carrot is abundant in temperate regions across the globe and is widely distributed across much of the United States where it is often found along roadsides, abandoned fields, and pastures. Wild carrot is the progenitor of the cultivated carrot, *D. carota* subsp. *sativus*, and the two subspecies are sexually compatible [24–26]. The cultivated carrot was likely domesticated in Central Asia roughly 1,100 yr ago [27,28] and is grown worldwide from both open pollinated and hybrid seed. Currently, no genetically engineered carrot products are available though much work has been carried out, see [29]; however, the carrot system is a good model for studying patterns of gene flow between crop-wild complexes. In many carrot-producing regions throughout the world, wild carrot populations can be found growing in close proximity to cultivated carrot fields [30,31]. Observations of wild carrot growing *within* cultivated carrot fields have also been reported [25,26,32]. In the eastern, midwestern, and northwestern United States, Queen Anne's Lace populations and crop carrot fields often occur in close proximity where gene flow between wild and crop populations of carrot is likely. Given the concerns for gene flow in both directions, we sought to investigate levels of crop-wild gene flow in two regions of carrot production within the United States utilizing population genetic methods.

In plants, gene flow among populations can occur through both pollen and seed owing to the differential transmission of nuclear and organellar genomes [33–36]. Nuclear genomes are biparentally inherited; thus, genetic material moves through both seed and pollen. Organellar genomes (mitochondrial and chloroplast) however, usually only move via pollen or seed because of uniparental inheritance. In flowering plants, uniparental inheritance is typically maternal for both the mitochondrial and chloroplast genomes. This differential mode of inheritance can lead to varying levels of gene flow for pollen versus seed since pollen generally moves greater distances than seed. In an extensive review, Petit et al. [37] found that when compared to biparentally inherited genomes, those that are maternally inherited experience more population subdivision—usually an order of magnitude less in terms of gene flow. Cytoplasmic male sterility is a maternally-inherited trait that is widely used in carrot breeding with molecular marker evidence supporting a maternal inheritance model, but with relatively few markers [31,38]. Similarly, maternal inheritance was observed for a plastid polymorphism [39], but few plastid molecular markers have been evaluated in carrot. So for both organellar

genomes, predominant, but not exclusive, maternal inheritance is documented. However, previous work on a limited sample of wild carrot populations in the eastern United States demonstrated lower than expected levels of population sub-structuring of mitochondrial genomes (i.e., indicating the potential for greater gene flow than expected for organellar genes) [40]. We therefore wished to compare the potential for seed and pollen to differentially contribute to gene flow in the current study by comparing nuclear and cytoplasmic (chloroplast) gene markers.

Given strict uniparental inheritance of organelles (maternal in the case of angiosperms), the comparison of gene flow in seed versus pollen is fairly straightforward. However, it is becoming increasingly evident that the rule of maternal inheritance in flowering plants is not always the case (see *Tansley Review* by [41]). In fact, in most taxonomic groups where this has been examined, some degree of non-strict-maternal inheritance, often termed *paternal leakage*, has been observed. From fish to fungi, natural and experimental studies have demonstrated DNA in the offspring originating from the paternal donor (reviewed recently in [42]). In wild carrot, indirect evidence of paternal leakage has been demonstrated through the observation of offspring that are heteroplasmic containing a mixture of at least two types of mitochondrial or chloroplast genomes [40,43]. Paternal leakage of organellar genomes should lead to less population subdivision than would typically be expected since organellar genes would move through both seed and pollen. Values of population sub-structuring (e.g., F_{ST} measures) would tend to approach those of nuclear genomes depending on the degree of paternal leakage. Moreover, because the two genomes (mitochondrial and chloroplast) need not leak together, values of population subdivision may be different for gene markers of the two genomes [41]. Bearing this in mind, we also wished to evaluate the results of gene flow in nuclear and chloroplast genomes in the context of any potential paternal leakage that might occur between crop and wild carrot populations.

Recently several authors have suggested additional methods (beyond F_{ST} and related measures) for assessing gene flow among plant populations [44–46] and in particular between crop and wild species populations [47]. These authors suggest using metrics, such as those based on Shannon information theory (i.e., from Ecology: Shannon diversity indices) [48], to investigate allele frequency variation within and among populations. Such indices have traditionally been employed in ecological studies to evaluate species diversity and evenness. One benefit of using Shannon diversity indices is that these measures account for allele identity along with frequency. Moreover, Campbell et al. [47] suggest comparing both total and multilocus measures when assessing the potential for gene flow between crop and wild populations. We therefore calculate both traditional population genetic metrics and ecological Shannon diversity indices and make comparisons at both the total and multilocus levels in the present study.

The overall goals of this study are to 1) investigate patterns of gene flow between crop and wild carrot across two regions of the United States using both nuclear and chloroplast DNA markers, 2) investigate evidence for chloroplast heteroplasmy in samples of both crop and wild carrot, and 3) compare conventional population genetic metrics of gene flow and genetic diversity to those that make use of species diversity indices of Shannon information.

Materials and Methods

Plant sampling strategy and genomic DNA isolation

The sampling strategy for wild carrot consisted of collecting leaf material or seeds (subsequently grown in the University of Memphis Greenhouses to obtain leaf material) from populations that varied from less than 1 km from a crop carrot field to more than 2.5 km from a crop carrot field. Carrot is a primarily outcrossing, insect pollinated species that is also capable

of self-fertilization [49]; seed can be dispersed by both wind and animals [50,51]. This strategy was carried out in two regions of the United States: Nantucket Island (Nantucket County) in Massachusetts in the eastern United States where carrot is grown as a root crop and on the Olympic Peninsula in the Sequim-Dungeness Valley (Clallam County) in Washington State in the western United States where crop carrot is grown for seed production. Three populations (Cliff Rd, Polpis Rd, Tuckernuck: 45 individuals) located away from crop fields and two populations (Bartlett Farms, Moors End Farm: 84 individuals) located nearby to crop fields of wild carrot were sampled from Nantucket in Massachusetts. Three populations (Hemlock, Kendall, Medsker: 43 individuals) located away from crop fields and four populations (Eberle, Fasola, Fencebird, Prince: 93 individuals) located near crop fields of wild carrot were sampled from the Olympic Peninsula/Sequim-Dungeness Valley of Washington State. The majority of leaves and seeds were collected from publicly available land, no specific permissions or permits were required for these locations, and this study did not involve endangered or protected species. Two populations were collected from private land and specific permission was obtained by JRM from the landowners for collection. Note that germination rates of *Daucus* seed were generally high (> 90%), and seeds were allowed to grow for up to 8 weeks before tissue was collected for DNA purification.

The four carrot cultivars (total 19 individuals) that are grown as a root crop on Nantucket were obtained from Johnny's Seed Company and Harris Seed Company (information kindly provided by the farmers on Nantucket). These varieties included 'Purple Haze F1', 'Nelson F1', 'Nantindo F1'and open-pollinated 'Scarlet Nantes'. It should be noted that even though carrot is cultivated for the root on Nantucket, farmers of both locations studied here report early bolters (plants flowering in their carrot fields in the first season) and plants occasionally flowering in their fields after harvest the following season (presumably missed during harvest the previous season). The three open-pollinated cultivars (total 49 individuals) grown in the Olympic Peninsula/Sequim-Dungeness Valley are self-selected lines grown by the farms: OP-1, OP-2, OP-3 (seed kindly provided by the farmers of the Olympic Peninsula). Since crop lines are continuously rotated (pers. comm. farmers of Nantucket and Olympic Peninsula/Sequim-Dungeness Valley), these lines were grouped into one population (per region) for the analyses presented here. Also note that carrot growers who grew a crop from the seed sources above, in both Nantucket and the Olympic Peninsula, reported no incidence of wild carrot outcrosses in their seed stocks.

Genomic DNA was isolated from a 2 cm portion of leaf tissue obtained from individual leaves using the SQ Omega Biotek Plant Kit (Atlanta, GA, USA) with the addition of 1% PVP and 1% Ascorbic Acid. DNA was quantified and quality assayed using a Nanodrop spectrophotometer (Thermo Scientific, MA). If DNA quality were low, samples were purified using an Omega Biotek E.Z.N.A. Cycle Pure Kit following the manufacturers protocol.

Chloroplast DNA length variation and nuclear microsatellites

Length variation of the region between the trnG and trnS genes (S/G region) of the chloroplast genome was used as a chloroplast genetic marker. PCR amplification of this region was performed using the 'S' and 'G' primers from Hamilton [52]. The PCR was carried out in a total reaction volume of 15 μl, consisting of 1.5 μl of 10X PCR buffer (100mM KCl, 100mM Tris HCl (pH9.0), 80mM (NH4)2SO4, and 1.0% Triton X-100), 0.35 μl MgCl2 (25mM), 0.2 μl dNTPs (each at 20mM), 0.2 μl forward primer (5μM), 0.2 μl reverse primer (20μM), 0.2 μl fluorescently labeled M13 primer (following the methods of [53] as in [54]) and (10μM), 0.5 μl of Taq DNA polymerase, and 1 μl template DNA (20 ng/μl). The PCR conditions were: 3 min at 95°C; 10 cycles of 30 s at 94°C, 30 s at 65°C and 45 s at 72°C, annealing temperature decreasing

to 55°C by 1°C per cycle, followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, 45 s at 72°C, followed by 10 min at 72°C. The PCR amplicons were diluted 1:10 and visualized using an ABI 3130xl DNA sequencer (Applied Biosystems, Foster City, CA) with GeneScan 500 LIZ dye Size Standard (Applied Biosystems) included in each lane to allow for accurate fragment size determination. Alleles were scored using the software package GeneMarker v. 2.6.3 (SoftGenetics, State College, PA). Heteroplasmic individuals were characterized on the basis of more than one allele signal for a single individual following the methods of Ellis et al. [55].

A total of 15 nuclear Simple Sequence Repeat (SSRs or microsatellite) markers previously developed by [56] were used in this study: GSSRs 3, 4, 6, 7, 9, 11, 16, 24, 31, 35, 57, 65, 85, 107, and 111. These markers are scattered throughout the genome and include all nine linkage groups. Methods for PCR amplification and visualization follow as above for chloroplast DNA with a few adjustments in amounts of reagents (see S1 Table for details).

Population genetic analyses

Measures of genetic diversity and genetic distance were assayed separately for the two regions studied here using the software GenAlEx v. 6.5 [57] for the two genomes. Given that heteroplasmy exists in wild populations of Queen Anne's Lace and the mixture is usually skewed toward one haplotype/allele [43], we chose to score chloroplast markers on the basis of the major haplotype (essentially ignoring heteroplasmy in measures of genetic diversity and gene flow here). Nei's [58] unbiased nuclear gene diversity uHe, chloroplast unbiased diversity cpuH, and the Shannon information index H were calculated for each population. Population structure was assessed using analysis of molecular variation (AMOVA; [59]), as implemented in GenAlEx, to hierarchically partition genetic variation and estimate Wright's F_{ST} [60]. Statistical significance (i.e., H_0 = no genetic differentiation among the populations) was determined by performing 1000 permutations. Further, genetic distance amongst individuals was investigated graphically using principal coordinates (PCO) analysis. For this, a standard genetic distance matrix [58] was constructed based on the nuclear multi-locus genotypes, the resulting matrix was used for the PCO analysis, and the first two principal coordinates were graphed in 2-dimensional space.

Population structure based on nuclear loci in each region ("Nantucket" and "Olympic") was also investigated using the Bayesian, model-based clustering algorithm implemented in the software package *structure* [61]. For these analyses, individuals were assigned to K population genetic clusters based on their nuclear multilocus genotypes. The software, *structure*, assembles clusters by minimizing intra-cluster Hardy–Weinberg and linkage disequilibrium and, for each individual, the proportion of membership in a given cluster is estimated. Our analysis here did not use prior population information (i.e., USEPOPINFO was turned off). For each analysis, K = 1–10 population genetic clusters were evaluated with 5 runs per K value. For each run, the initial burn-in period was set to 200,000 with 500,000 MCMC iterations. The optimal number of K clusters was estimated using the delta K method of Evanno et al. [62] as implemented in Structure Harvester on the web [63]. When an optimal K was determined for either region, membership coefficients in each cluster were aligned across replicate runs using *CLUMPP* [64] using the *Greedy* algorithm. Finally, the cluster membership was visualized using the software *distruct* [65]. Pie charts of structure results were also graphically visualized using ArcGIS (Release 10.2.2. Redlands, CA: Environmental Systems Research Institute).

In order to estimate contemporary rates of gene flow among populations within regions and as a complement to our historical methods using traditional population genetic analyses, we used the software BayesAss 3.0.4 [66]. As advised by the authors, we varied seed numbers and monitored consecutive runs in order to diagnose convergence. Also, following the authors'

suggestion, delta values were adjusted so that the accepted numbers of changes were between 20–60%. MCMC runs were performed with 10,000,000 iterations after a burn-in of 1,000,000 and a sampling frequency of 1,000. The migration rates for each population, including the standard deviation of the marginal posterior distribution for each migration rate estimate, were recorded.

Comparison of traditional population genetic measures versus ecological measures

Traditional population genetic measures of Nei's [58] unbiased gene diversity (uHe) and Wright's F_{ST} [60] were compared to those derived from Shannon information theory (i.e., from Ecology: Shannon diversity indices) [44–46]. The population genetic measures uHe and F_{ST} both measure aspects of genetic diversity and differentiation within and among populations. The measure uHe is the expected heterozygosity, or gene diversity, corrected for sample size for a given population and ranges from 0 to 1 with higher values indicating greater population genetic diversity. The measure F_{ST} describes the degree of genetic differentiation among populations and also ranges from 0 to 1; when F_{ST} is lower, populations do not differ greatly in terms of allelic diversity, and when F_{ST} is higher, populations are described as more genetically different from one another. The Shannon measures, Shannon's H and S_{HUA} are analogous to the population genetic measures uHe and F_{ST} respectively: Shannon's H measures levels of genetic diversity within a population but is not bounded by 1; S_{HUA} (also not bounded by 1) similarly measures the degree of genetic differentiation amongst sampled populations. For measures of uHe and Shannon's H and F_{ST} and S_{HUA}, populations were grouped according to their proximity to crop fields (see description above of collection strategy). A two-factor ANOVA was performed in JMP (v. 9) using crop proximity and locus as factors to investigate significance among diversity measures. A sign test (in JMP) was also performed to assay significance.

Results

Chloroplast DNA

In Nantucket (the crop production sampling area), three haplotypes were recorded across the six populations (no private haplotypes). All populations except Tuckernuck were polymorphic for the chloroplast marker, and unbiased diversity ranged from 0 at Tuckernuck to 0.56 at Polpis Rd demonstrating substantial haplotypic variation across the island (Table 1). The F_{ST} value as estimated from AMOVA was 0.43 (p < 0.001) with haplotypic frequency showing some degree of structuring across populations (Fig 1). Two wild individuals (out of 129) from different populations showed evidence of chloroplast heteroplasmy, and four crop samples (out of 19) showed evidence of chloroplast heteroplasmy.

In the Olympic Peninsula (the seed production sampling area), six haplotypes were recorded across the eight populations (no private haplotypes). All populations were polymorphic for the chloroplast marker and unbiased diversity ranged from 0.31 in the crops to 0.69 in the Medsker Rd population demonstrating substantial haplotypic variation across the region (Table 1). The F_{ST} value as estimated from AMOVA was lower than on Nantucket: 0.29 (p < 0.001) with haplotypic frequency showing some degree of structuring across populations with populations on the northern end of the valley harboring a different haplotype frequency than populations farther south and farther from crop fields (Fig 2). Thirty-four wild individuals (out of 136) from three populations showed evidence of chloroplast heteroplasmy, and one crop sample (out of 49) showed evidence of chloroplast heteroplasmy.

Table 1. Carrot population information and genetic diversity measures for nuclear and chloroplast loci.

Region	Population	Proximity	Number	uHe (SE)	Shannon H	cpuH
Nantucket	Crop	—	19	0.59 (0.04)	1.09 (0.10)	0.50
Nantucket	Cliff Rd	Away	19	0.60 (0.06)	1.16 (0.12)	0.53
Nantucket	Polpis Rd	Away	10	0.64 (0.05)	1.13 (0.11)	0.56
Nantucket	Tuckernuck	Away	16	0.58 (0.03)	0.99 (0.06)	0.00
Nantucket	Bart Farm	Near	37	0.73 (0.03)	1.56 (0.11)	0.47
Nantucket	Moors Farm	Near	47	0.72 (0.04)	1.60 (0.12)	0.04
Olympic	Crop	—	49	0.42 (0.06)	0.77 (0.12)	0.31
Olympic	Hemlock	Away	14	0.80 (0.03)	1.71 (0.09)	0.62
Olympic	Kendall	Away	16	0.67 (0.03)	1.25 (0.09)	0.46
Olympic	Medsker	Away	13	0.78 (0.03)	1.64 (0.13)	0.69
Olympic	Eberle	Near	36	0.76 (0.03)	1.60 (0.10)	0.48
Olympic	Fasola	Near	11	0.62 (0.06)	1.16 (0.12)	0.33
Olympic	Fencebird	Near	31	0.70 (0.03)	1.44 (0.08)	0.40
Olympic	Prince	Near	15	0.69 (0.03)	1.30 (0.07)	0.25

uHe = Nei's (1978) unbiased nuclear gene diversity ± standard error (SE). Shannon H = Shannon Information Index for nuclear loci (Peakall and Smouse 2012) ± standard error (SE). cpuH = chloroplast unbiased diversity (Peakall and Smouse 2012).

doi:10.1371/journal.pone.0161971.t001

Nuclear microsatellites

In Nantucket (the crop production sampling area), all fifteen nuclear SSRs were polymorphic in all populations, and unbiased multilocus gene diversity ranged from 0.58 on Tuckernuck to 0.73 at Bartlett Farms (Table 1). Multilocus Shannon information index values (Shannon H) varied from 0.99 on Tuckernuck to 1.60 at Moors End Farm (Table 1). Single locus values were also calculated for both uHe and Shannon H and compared using a sign test and proximity and locus two-factor ANOVA. For both metrics, wild populations in close proximity to crop fields showed overall *higher* single locus values of uHe and Shannon H, though the Shannon H measures appeared more sensitive to these differences (uHe: 11/15 loci higher in near versus away populations; sign test $p = 0.059$; proximity factor ANOVA, $F = 9.01$, $p = 0.0095$ and Shannon H: 14/15 loci higher in near versus away populations; sign test $p < 0.0001$; proximity factor ANOVA, $F = 28.44$, $p < 0.0001$; Fig 3).

In terms of population structure, the multilocus F_{ST} value as estimated from AMOVA was 0.14 ($p < 0.001$) with paired population F_{ST} values ranging from 0.11 (Polpis Rd to Cliff Rd—both away populations) to 0.30 (Crops to Cliff Rd) (S1 Fig). The 2-dimensional graphical representation of the populations (Fig 4) demonstrates a likely high degree of population connectivity across the island with some structure but substantial overlap amongst populations. The DeltaK method of Evanno et al. [62] as applied to the Bayesian clustering method *structure* [61] provided support for the presence of four genetically distinct clusters in Nantucket (Fig 1 and S2 Fig). These clusters roughly corresponded to the crop individuals, away proximal populations and then the two separate wild populations located near carrot farms; yet some degree of genotypic/allelic overlap can be seen among individuals (S2 Fig).

When F_{ST} was calculated on the grouping of populations based on proximity in Nantucket (the crop production sampling area), the near to crop populations versus far from crop populations demonstrated slightly lower values: away wilds to crops $F_{ST} = 0.23$, near wilds to crops $F_{ST} = 0.21$. A single locus approach based on proximity also failed to demonstrate any significant differences in away versus near measures of F_{ST}, proximity and locus two-factor ANOVA $p = $ NS. However, when Shannon diversity indices were used, the difference between distance
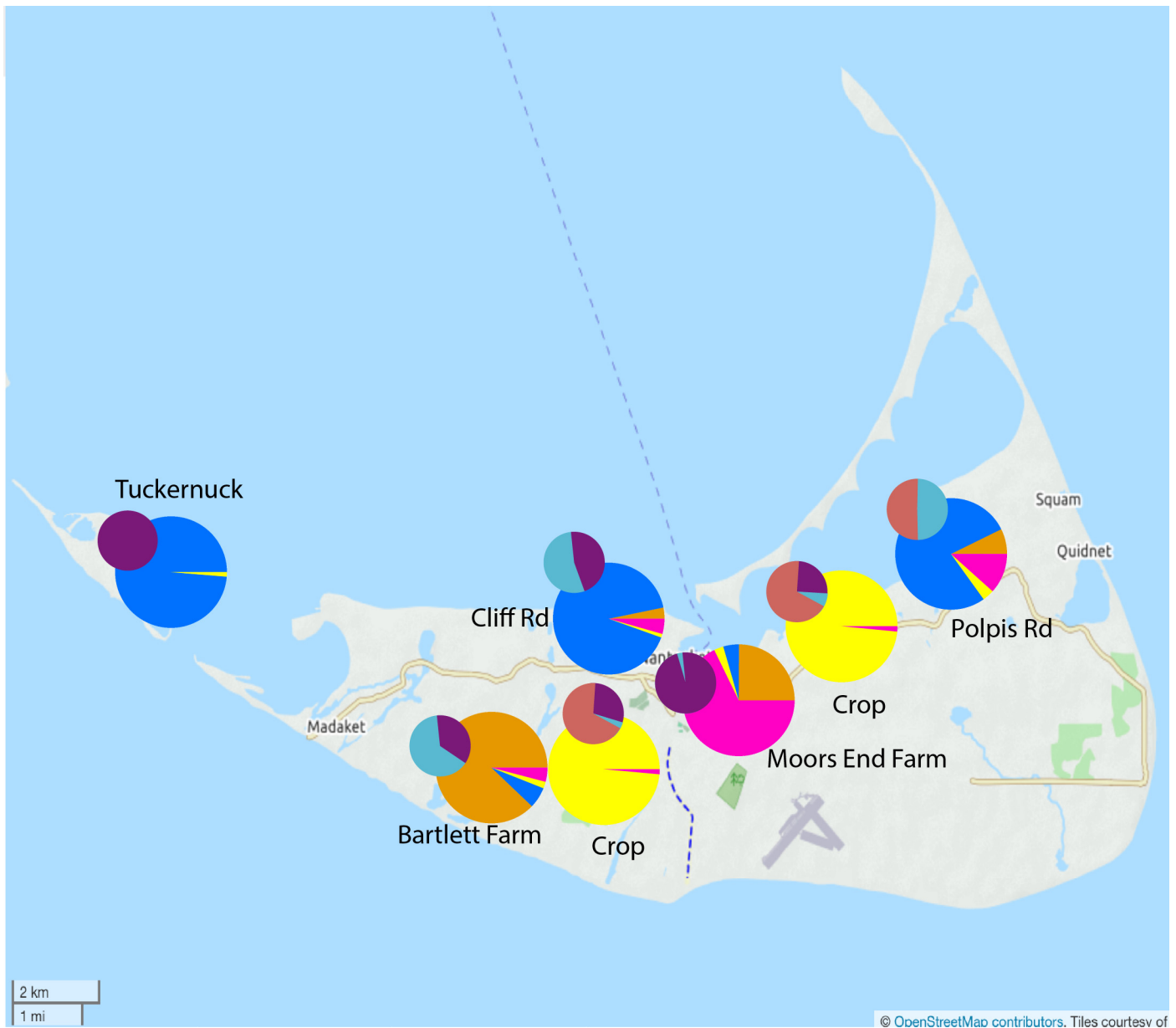
Fig 1. Nantucket *structure* analysis. Results of *structure* analysis for K = 4 clusters (large pie charts) and chloroplast haplotypic diversity (small pie charts) for Nantucket Island.

measures was greater: away wilds to crops Shannon Hua = 0.21; near wilds to crops Shannon Hua = 0.12, and the single locus approach demonstrated that the Shannon Hua values were lower in near versus away populations, proximity and locus two-factor ANOVA, F = 25.12 p = 0.0002. Contemporary estimates of migration using the software program BayesAss yielded estimates of migration rates (range: 0.068% to 21%) that were relatively similar across populations for Nantucket. The largest rate was the Cliff Rd population as an immigration source for the Polpis Rd population—both populations located away from crop fields (S2 Table).

In the Olympic Peninsula (the seed production sampling area), fourteen nuclear SSRs (GSSR9 yielded unreliable amplification and was removed from further analyses) were polymorphic in all populations, and unbiased multilocus gene diversity ranged from 0.42 in the
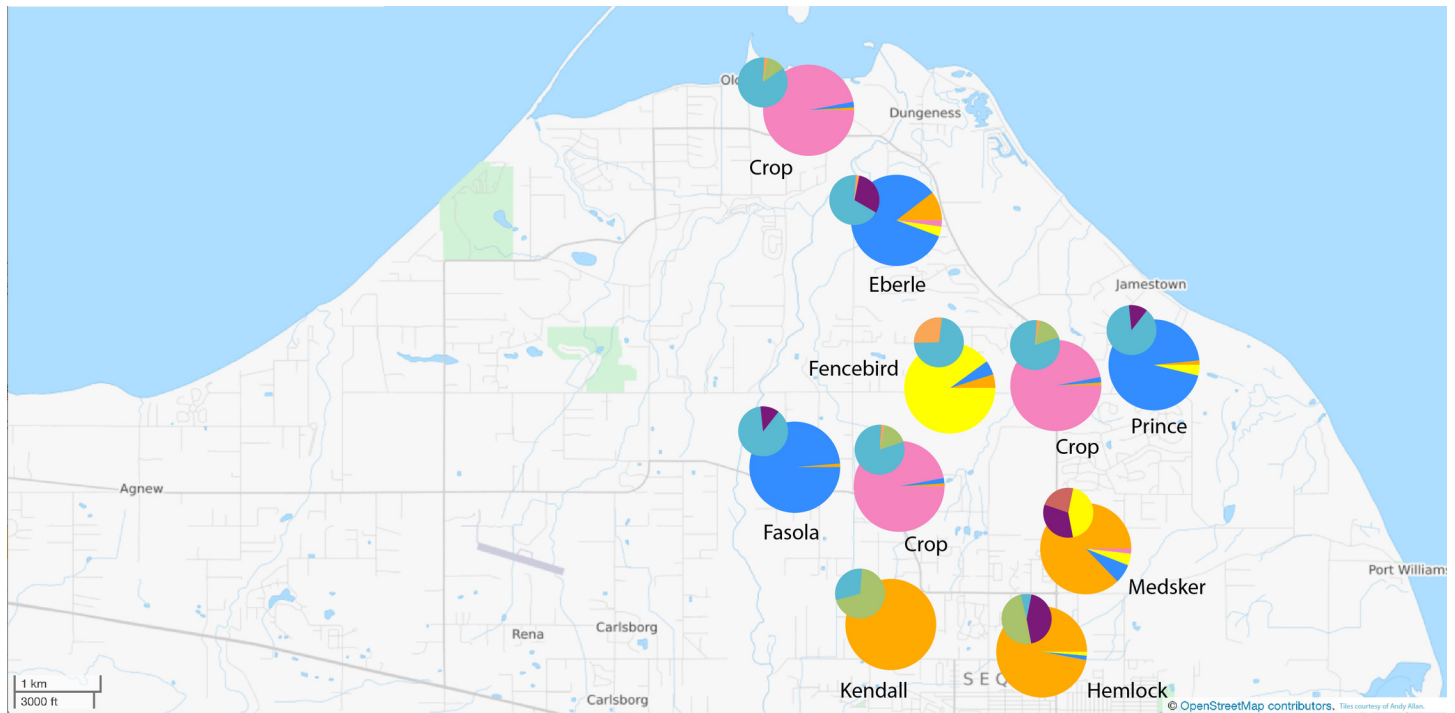
Fig 2. Olympic *structure* analysis. Results of *structure* analysis for K = 4 clusters (large pie charts) and chloroplast haplotypic diversity (small pie charts) for Olympic Peninsula/Sequim-Dungeness Valley.

doi:10.1371/journal.pone.0161971.g002

crops to 0.80 at Hemlock Rd population ([Table 1](#)). Multilocus Shannon information index values (Shannon H) varied from 0.77 in the crops to 1.71 at Hemlock Rd population ([Table 1](#)). Single locus values were also calculated for both uHe and Shannon H and compared using a sign test and proximity and locus two-factor ANOVA. Different from Nantucket, for both metrics, wild populations in close proximity to crop fields showed overall *lower* single locus values of uHe and Shannon H, and the Shannon H measures appeared more sensitive to these differences (uHe: 11/15 loci *lower* in near versus away populations; sign test p = NS; proximity factor ANOVA p = NS and Shannon H: 11/15 loci *lower* in near versus away populations; sign test p = NS; ANOVA, proximity factor F = 4.76, p = 0.027; [Fig 5](#)).

In terms of population structure, the multilocus F_{ST} value as estimated from AMOVA was 0.17 (p < 0.001) with paired population F_{ST} values ranging from 0.049 (Hemlock to Medsker —both away populations) to 0.48 (Crops to Fasola—a near population) ([S3 Fig](#)). The 2-dimensional graphical representation of the populations ([Fig 6](#)) demonstrates a substantial amount of population connectivity across the valley with some structure but substantial overlap amongst populations. The DeltaK method of Evanno et al. [[62](#)] as applied to the Bayesian clustering method *structure* [[61](#)] provided support for the presence of four genetically distinct clusters in Olympic ([Fig 2](#) and [S4 Fig](#)). These clusters corresponded to the crop individuals, away populations, and near to crop field populations. Still some degree of genotypic/allelic overlap within individuals can be seen ([S4 Fig](#)).

When F_{ST} was calculated on the grouping of populations based on proximity in Olympic (the seed production sampling area), the near to crop populations versus away to crop populations demonstrated slightly lower values: away wilds to crops F_{ST} = 0.27, near wilds to crops F_{ST} = 0.22. The single locus approach based on proximity also failed to demonstrate any significant differences in away versus near measures of F_{ST}, proximity and locus two-factor ANOVA
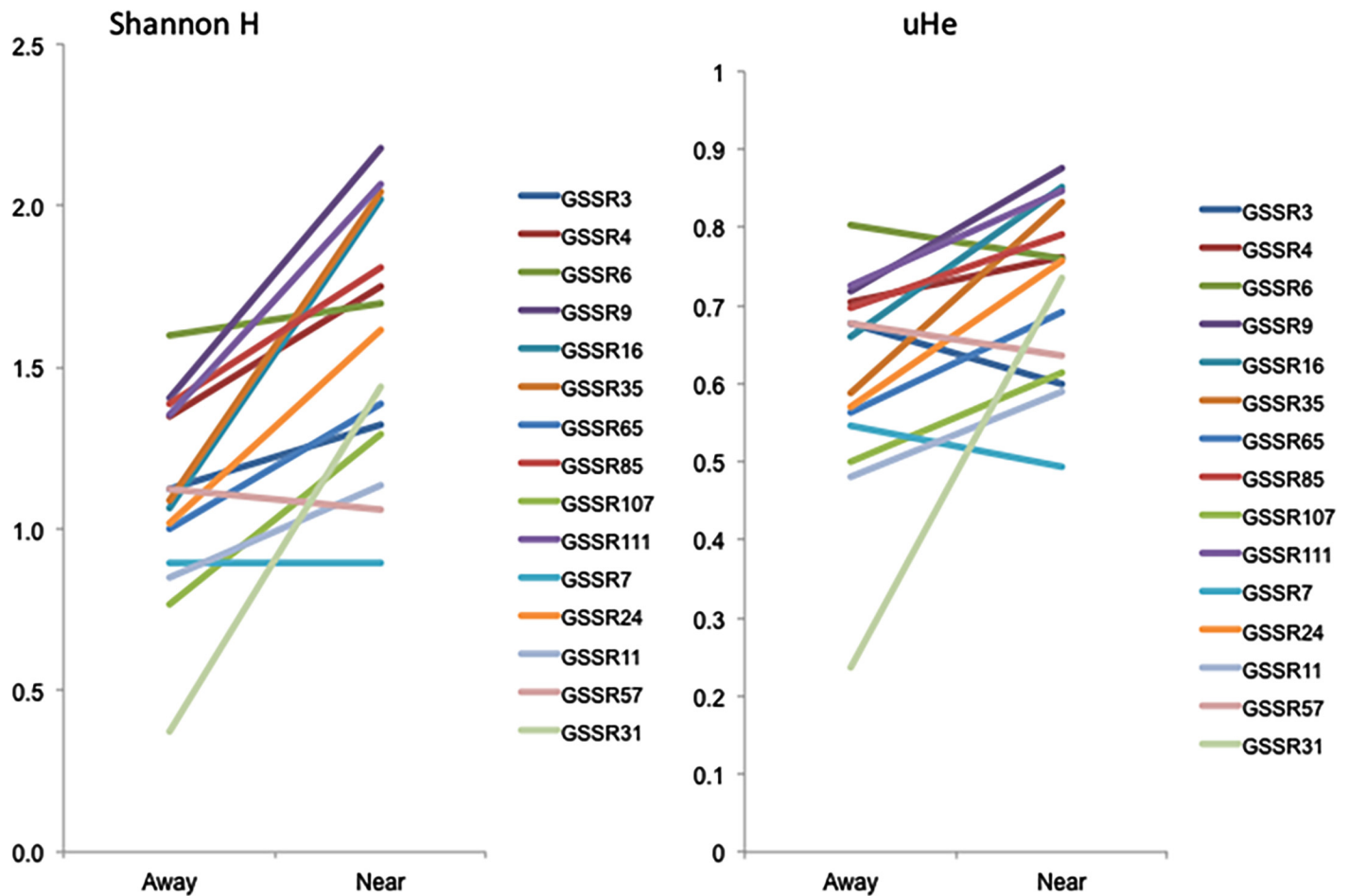
Fig 3. Nantucket Shannon H and gene diversity. Comparison of Shannon H and unbiased gene diversity on a per locus basis for away versus near wild carrot populations for Nantucket.

(p = NS). When Shannon diversity indices were used, the away populations also showed greater divergence than the near populations but not as large as the difference in Nantucket: away wilds to crops Shannon Hua = 0.22; near wilds to crops Shannon Hua = 0.17; yet still the single locus approach demonstrated that the divergence measures of Shannon Hua were lower in near versus away populations, proximity and locus two-factor ANOVA, F = 20.18 p = 0.0006. Contemporary estimates of migration using the software program BayesAss yielded estimates of migration rates (range: 0.058% to 21%) that were relatively similar across populations for Olympic. The largest rate was the Eberle population as an immigration source for the Fasola population—both populations located near crop fields (S2 Table). Raw data allele scores for nuclear and chloroplast loci are located in S3 Table.

Discussion

Patterns of nuclear gene flow and comparison of methods

A major goal of this study was to investigate the evidence for gene flow among populations of carrot with particular interest in levels and patterns of gene flow occurring amongst crop and wild populations. Thus, we sampled from wild carrot locales that were within 1km of a crop field and locations that were more than 2.5km away from crop fields in two regions where
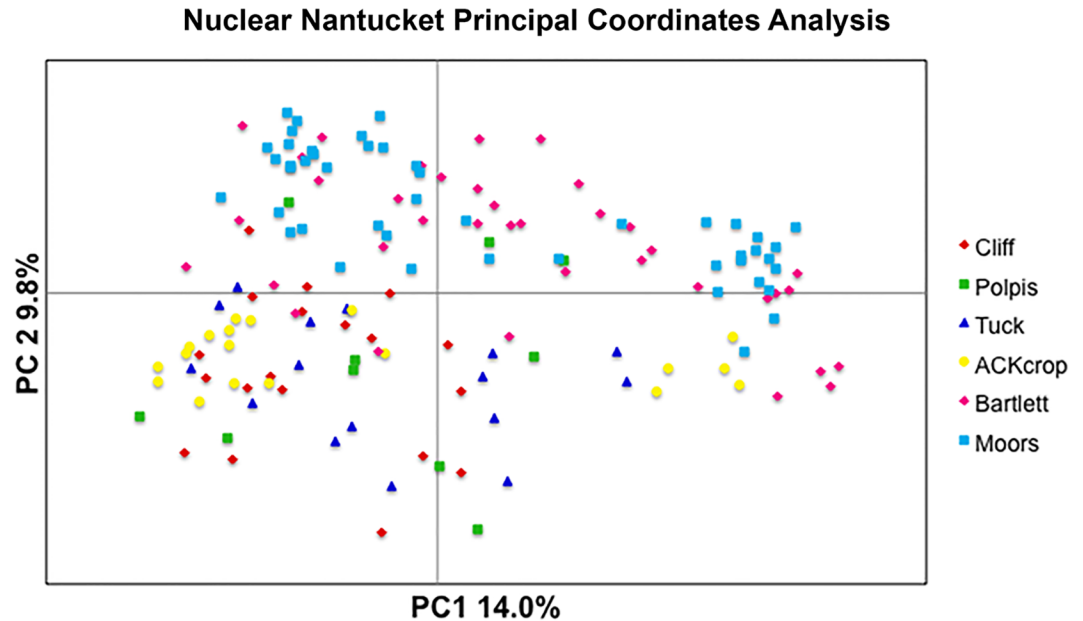
Nuclear Nantucket Principal Coordinates Analysis

Fig 4. Nantucket Principal Coordinates Analysis. Results of the nuclear Nantucket Principal Coordinates Analysis graphically represented in 2-dimensional space.

carrot is grown in the United States (recognizing that this represents a reasonable estimate of historical levels of gene flow). Using both methods that detect historical levels of gene flow (F_{ST}) as well as methods more traditionally used for assessing ecological species diversity (i.e., Shannon diversity indices), the results presented here demonstrate a significant level of gene flow among populations both on Nantucket and in the Olympic Peninsula/Sequim-Dungeness Valley. Our results show that while populations do exhibit genetic structuring, some degree of population connectivity among crop and wild carrot populations certainly occurs. Moreover, gene flow appears to occur at a significant rate with respect to *both* nuclear and cytoplasmic genomes. The measures of nuclear F_{ST} reported here (Nantucket F_{ST} = 0.14; Olympic F_{ST} = 0.17) are on par with those of other angiosperm species as reported by Petit et al. [37] where the biparental mean ± standard error/SE from 77 studies was 0.184 ± 0.002. Additionally, a recent study including crop and a global sampling of wild carrot individuals reported moderate levels of genetic differentiation (0.19; [27]).

When we compared measures of genetic differentiation here grouping populations based on their proximity to crop fields, we found that in both regions (Nantucket and Olympic) populations that were closer to crop fields harbored lower levels of differentiation as compared to populations that were farther. This finding held when using both measures of F_{ST} and Shannon diversity indices; however, the Shannon measures were more sensitive to these differences (stronger statistical support). This result aligns with recent suggestions by several authors to make use of both types of measures when investigating population structure and gene flow [44–46] especially between crop and wild species populations [47]. Moreover, Campbell et al. [47] suggest taking a single locus based approach (analyzing single loci separately along with total multilocus measures). When we followed this approach, we found that accounting for single locus effects yielded more sensitivity and stronger statistical support for differentiation and levels of genetic diversity for Shannon diversity indices but not for conventional measures of F_{ST} (similar to the findings of [47]).

An additional finding of Campbell et al. [47] was that putative hybrid populations (between wild plants and their outcrossing crop relatives) demonstrated increased genetic diversity
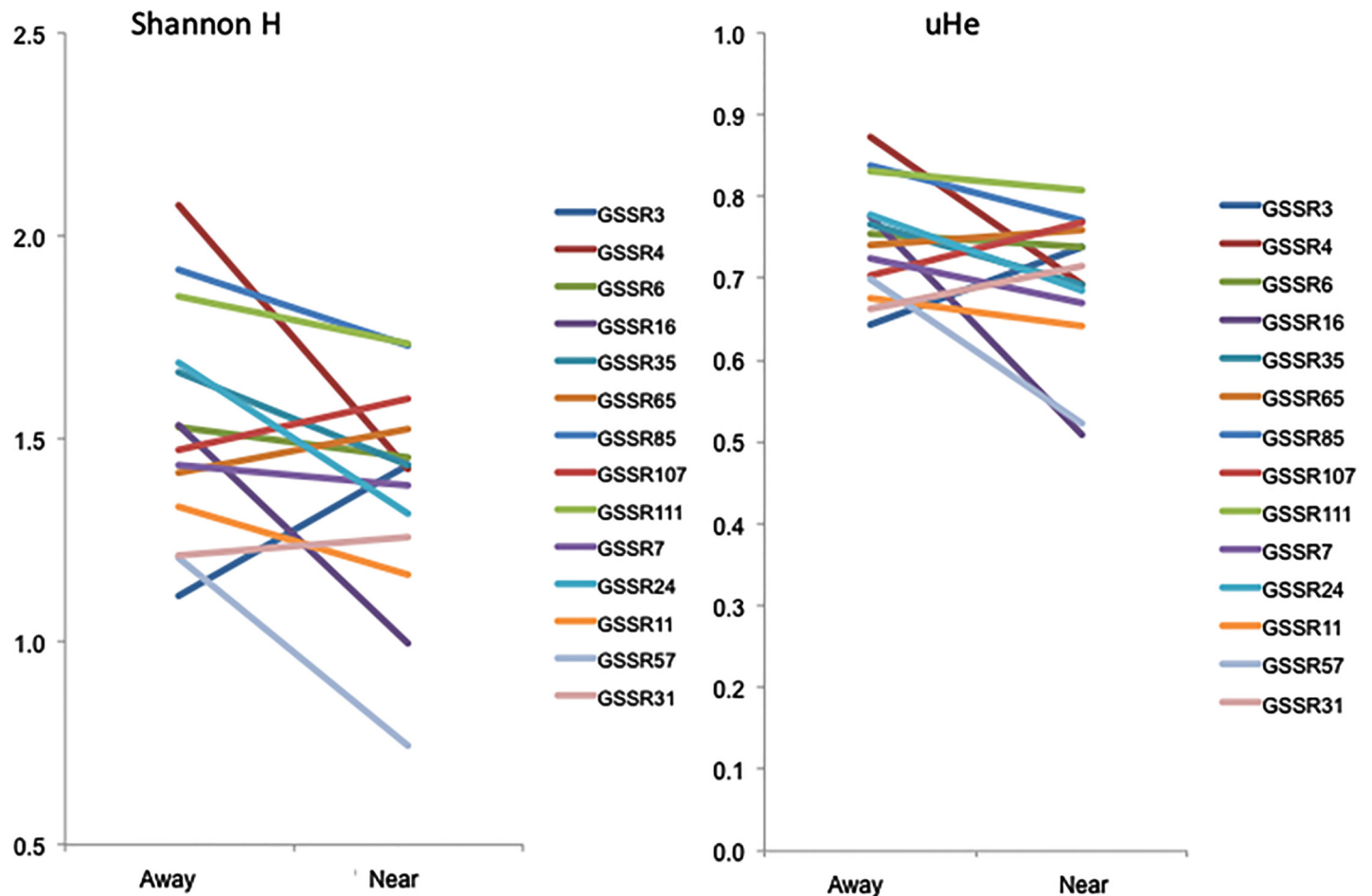
Fig 5. Olympic Shannon H and gene diversity. Comparison of Shannon H and unbiased gene diversity on a per locus basis for away versus near wild carrot populations for Olympic.

(using ecological measures) when compared to non-hybrid populations. Given this, one might predict that our populations in close proximity to crop fields would show evidence of increased genetic variation. Interestingly, this increased variation was the case for the Nantucket region but the opposite trend was found in the Olympic region (with Shannon measures again being more sensitive than those of unbiased gene diversity). We can currently only speculate as to why these differences between regions might occur: one possibility is the difference in management strategies of wild carrot populations between the two regions. One possibility relates to the levels of genetic variation in the crops: Nantucket crops were more genetically diverse in all measures when compared to crops on the Olympic Peninsula. One finding to note here is that given the Nantucket region is a root field site, we might have expected little to no gene flow at this location between crop and wild populations; however, our study did find population connectivity at the genetic level though this varied depending on the marker type used (measures of population structure at nuclear loci were similar in both regions while chloroplast population structure was higher on Nantucket).

In both regions, wild carrot is very common and on Nantucket it is regarded as an attractive wild flower that residents cultivate in their own home gardens. The species is also tolerated or ignored on the farm properties where crop carrot is grown with farmers occasionally reporting early bolters and plants with intermediate root morphology (potential F_1 hybrids) providing
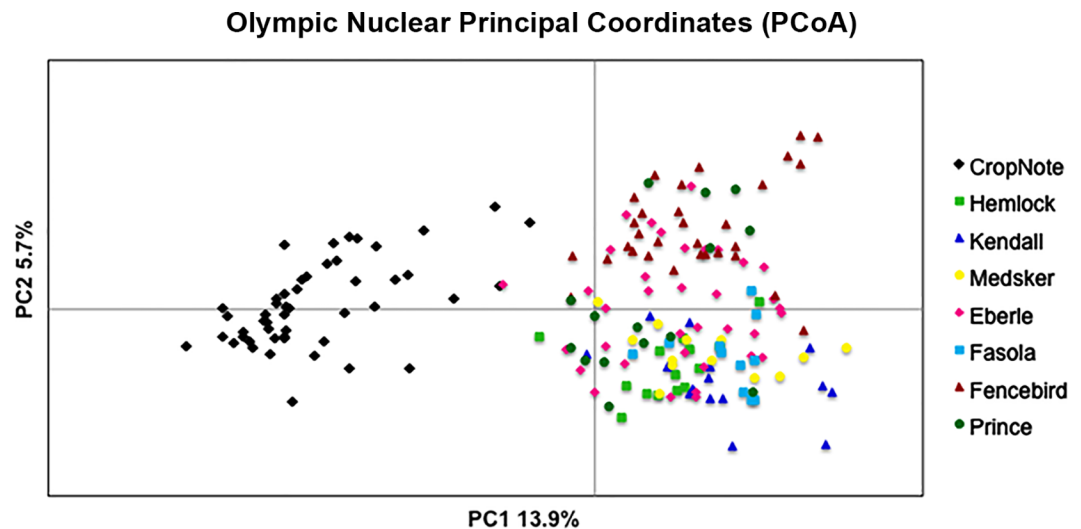
Olympic Nuclear Principal Coordinates (PCoA)

Fig 6. Olympic Principal Coordinates Analysis. Results of the nuclear Olympic Principal Coordinates Analysis graphically represented in 2-dimensional space.

doi:10.1371/journal.pone.0161971.g006

evidence for outcrossing with wild carrot when seed for that crop was produced. The situation is quite different in the Olympic Peninsula/Sequim-Dungeness Valley where commercial carrot seed production is an important priority of the region. Both the local farmers and the Clallam County Extension Office have active campaigns against wild carrot due to concerns of contamination of crop seed. "Wanted" posters can be found in the area stating "Queen Anne's Lace, Wanted Dead", and the community is advised to cut the umbels of the plants off before they set seed.

Another possibility is in the difference in carrot seed sources used by farmers in the two regions. Farmers in the Nantucket region are often using hybrid carrot seed and likely very little of that carrot seed has been produced locally. So any seed set in nearby wild carrot populations that results from pollination with cultivated carrot is relatively rare. And with minimal wild carrot control, those pollinations more likely involved wild carrots closer to crop production fields. In contrast, production of open-pollinated carrot seed is an ongoing practice by farmers in the Olympic Peninsula. While wild carrot control measures are applied in the Olympic Peninsula, the ongoing annual presence of pollen from cultivated carrots may well have resulted in gene flow to wild carrots from local seed production fields on a regular basis to the extent that wild carrots closer to production fields are more similar to locally cultivated carrots that more distant populations of wild carrots. Additional surveys of populations near and far from crop fields in both regions would be useful to see if this pattern holds with a broader sampling. Furthermore, an evaluation of specific alleles associated with domestication traits like root color, lateral root formation, and early bolting (floral initiation) in populations near to and far from crop production fields in future studies may reflect differential selection for those genomic regions relative to random genetic markers.

Patterns of chloroplast gene flow

While measures of nuclear genetic differentiation follow those of other angiosperm species, the measures of population structure based upon the chloroplast marker were much lower than typically found in angiosperm plant populations (Nantucket F_{ST} = 0.43; Olympic F_{ST} = 0.29). Petit et al. [37] summarized 124 angiosperm population genetic studies that surveyed

maternally inherited markers and found that the mean genetic differentiation was 0.637 (± 0.002 standard error). The median value was 0.646 with the 25th and 75th quartiles being 0.416 and 0.871 respectively. Our chloroplast differentiation values were either on the extreme low end (for Nantucket) or well outside the range of the quartiles (for Olympic). Other studies have found similar low values for carrot organellar population structure. Ronfort et al. [67] examined population structure in populations of wild carrot in France and found levels of mitochondrial genetic differentiation to be extremely low (0.08); while Mandel et al. [40] also found lower than expected values of mitochondrial population structure in natural populations of wild carrot in the Eastern United States (0.34).

Several explanations can be put forward for this finding of lower than expected cytoplasmic gene flow in our study. One possibility is that seeds of carrot are better dispersers (i.e., potential for greater gene flow) than many other angiosperm species such that the chloroplast genome shows significantly lower levels for differentiation as compared to other taxa [37]. Indeed, carrot seeds do tend to move substantial distances in the wild being dispersed by both animals and wind with wind possibly contributing more to dispersal than animals [50,51]. Thus, it appears that efficient dispersal of seeds may play a role in increasing population connectivity and decreasing population structure with respect to organellar genomes. Furthermore, prior to the development of large-scale commercial carrot seed production in regions of the world with little wild carrot and accompanying management systems put in place to minimize the occurrence of outcrosses (both in just the last 75 years), carrot seed was produced locally in Europe, Asia, and the Americas where wild carrot is known to have been widespread. This opened up the likelihood that cytoplasmic gene flow has been ongoing between wild and cultivated carrot throughout most of the history of the crop, and perhaps accounts for the reduced genetic differentiation observed today.

The action of metapopulation dynamics and frequent extinction and recolonization events may also play a role in determining population genetic structure in carrot populations. Ronfort et al. [67] point out that since wild carrot often inhabits newly disturbed areas and can be quickly overturned by later successional species, populations may experience some of the effects associated with extinction-recolonization dynamics. In metapopulations, both the number of colonists (starting new populations) and migration (among current populations) can influence gene flow and population genetic differentiation. When the number of colonists is high relative to among population migration rates, extinction and recolonization can enhance gene flow and diminish population genetic differentiation [68,69]. Given this, aspects of carrot population dynamics could play a role in determining the degree of differentiation especially if seeds are efficient dispersers to new unoccupied patches or habitats. It should also be noted that in both regions where we sampled these carrot populations, Queen Anne's Lace is virtually ubiquitous throughout Nantucket Island and the Olympic Peninsula/Sequim-Dungeness Valley (despite efforts to remove/control wild carrot here).

Another possibility is that the chloroplast genome is not transmitted strictly maternally but occasionally moves via pollen as well. Paternal leakage would allow chloroplast genes to move through both seed *and* pollen leading to higher levels of cpDNA gene flow and thus lower levels of population genetic differentiation. Occasional paternal inheritance of organellar DNA, sometimes termed "paternal leakage" has been documented in numerous plant species (nicely reviewed in [41]). Paternal inheritance of chloroplast DNA would tend to reduce measures of F_{ST} aligning them closer to F_{ST} values of the nuclear genome since the chloroplast genome could move in both seed and pollen. Paternal leakage of the chloroplast genome is also one factor that can lead to heteroplasmy within an individual (having a mixture of different organellar genes/genomes) making heteroplasmy an indirect indicator of paternal leakage in crosses [41]. Paternal inheritance of chloroplasts was previously reported in hybrid crosses between

different species of *Daucus* [70]. And recently, heteroplasmy of the mitochondrial genome has been documented in both crop [71] and wild carrot [43]. In the current study, we found evidence for heteroplasmy of the chloroplast genome in Nantucket and to a greater degree in Olympic Peninsula. Further supporting the potential for paternal leakage to contribute to chloroplast gene flow here is the finding of greater heteroplasmy in Olympic accompanied by a lower level of population genetic differentiation there. Further study including broader population sampling and crossing studies are needed to determine which, or which combination, of these possibilities contributes to the substantial degree of chloroplast gene flow in carrot populations.

Implications for transgene escape

Our findings have implications for crop-wild transgene escape especially that of organellar transgene escape. Two observations from the literature apply to this: 1) paternal leakage of otherwise maternally inherited cytoplasmic genomes is often associated with hybridization events [72–74]. For example, Boblenz et al. [70] found evidence for paternal inheritance of plastids in interspecific crosses among *Daucus* species. And 2) paternal transmission may also be enhanced in species that exhibit gynodioecious mating systems (consisting of both hermaphrodites and male-sterile females) as well as crop breeding programs that utilize cytoplasmic male sterility (CMS) like carrot [41,75]. The first observation that paternal leakage is enhanced in hybridization events suggests that divergence may play a role in permitting leakage in crosses such that it is more likely for leakage to occur in crosses between divergent populations (e.g., crop to wild). And secondly, species including crops that exhibit CMS may be particularly susceptible to paternal leakage of cytoplasmic DNA in their male-fertile lineages (note that CMS is essentially a breeding tool, and that male fertility is typically restored in seed sold to farmers), and thus cytoplasmic transgene escape. Therefore, studies which investigate heteroplasmy and paternal leakage in crop wild systems (especially those considering genetic modification) are certainly warranted. Finally, our study examined wild populations at distances less than, and greater than, 2.5km from crop fields. While we showed that gene flow is more abundant nearby to crop fields, gene flow likely occurs at greater distances as well. Certainly, consistent with previous findings [49] our study indicated that a minimum distance of 2.5km should be maintained, when possible, to limit the potential for gene flow with greater distances providing additional protection against gene flow to wild carrot populations.

Supporting Information

S1 Fig. Pairwise values of F_{ST} for Nantucket.
(TIF)

S2 Fig. Individual *structure* results for K = 4 clusters for Nantucket.
(TIF)

S3 Fig. Pairwise values of F_{ST} for Olympic.
(TIF)

S4 Fig. Individual *structure* results for K = 4 clusters for Olympic.
(TIF)

S1 Table. Details of PCR reactions for each locus.
(XLSX)

S2 Table. Migration estimates for each population for Nantucket and Olympic.
(XLSX)

S3 Table. Raw data microsatellite allele calls.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: JRM PWS.

Data curation: JRM AJR.

Formal analysis: JRM AJR MI PWS.

Funding acquisition: JRM.

Investigation: JRM AJR MI PWS.

Methodology: JRM AJR MI PWS.

Project administration: JRM.

Resources: JRM AJR MI PWS.

Supervision: JRM MI PWS.

Validation: JRM AJR.

Visualization: JRM AJR.

Writing – original draft: JRM AJR MI PWS.

Writing – review & editing: JRM AJR MI PWS.

References

1. Ellstrand N (2003) Dangerous Liaisons? When cultivated plants mate with their wild relatives. Baltimore, MD: Johns Hopkins.
2. Warwick SI, Stewart CN Jr (2005) Crops come from wild plants—how domestication, transgenes, and linkage together shape ferality. In: Gressel J, editor. Crop Ferality and Volunteerism. Boca Raton, FL: CRC Press. pp. 9–30.
3. Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annu Rev Ecol Syst 30: 539–563.
4. Stewart CN Jr, Halfhill MD, Warwick SI (2003) Transgene introgression from genetically modified crops to their wild relatives. Nat Rev Genet 4: 806–817. PMID: 14526376
5. den Nijs HCM, Bartsch D, Sweet J (2004) Introgression from Genetically Modified Plants into Wild Relatives. Wallingford: CABI.
6. Chapman MA, Burke JM (2006) Letting the gene out of the bottle: the population genetics of genetically modified crops. New Phytol 170: 429–443. PMID: 16626466
7. Kwit C, Moon HS, Warwick SI, Stewart CN Jr (2011) Transgene introgression in crop relatives: molecular evidence and mitigation strategies. Trends Biotechnol 29: 284–293. doi: 10.1016/j.tibtech.2011.02.003 PMID: 21388698

8. Schierenbeck KA, Ellstrand NC (2008) Hybridization and the evolution of invasiveness in plants and other organisms. Biol Invasions 11: 1093–1105.

9. Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, et al. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evol Appl 3: 494–504. doi: 10.1111/j. 1752-4571.2010.00140.x PMID: 25567942

10. Andersson MS, de Vicente MC (2010) Gene Flow Between Crops and Their Wild Relatives. Baltimore, MD: Johns Hopkins Univ. Press.

11. Ellstrand NC, Meirmans P, Rong J, Bartsch D, Ghosh A, de Jong TJ, et al. (2013) Introgression of crop alleles into wild or weedy populations. Annu Rev Ecol Evol Syst 44: 325–345.

12. De Wet JMJ, Harlan JR (1975) Weeds and Domesticates: Evolution in the Man-Made Habitat. Econ Bot 29: 99–107.

13. Milne RI, Abbott RJ (2000) Origin and evolution of invasive naturalized material of Rhododendron ponticum L. in the British isles. Mol Ecol 9: 541–556. PMID: 10792698

14. Nassar NMA (2002) Cassava, Manihot esculenta Crantz, genetic resources: origin of the crop, its evolution and relationships with wild relatives. Genet Mol Res 1: 298–305. PMID: 14963820

15. Snow AA, Pilson D, Rieseberg LH, Paulsen MJ, Pleskac N, Reagon MR, et al. (2003) A Bt transgene reduces herbivory and enahnces fecundity in wild sunflower. Ecol Appl 13: 279–286.

16. Dechaine JM, Burger JC, Chapman MA, Seiler GJ, Brunick R, Knapp SJ, et al. (2009) Fitness effects and genetic architecture of plant-herbivore interactions in sunflower crop-wild hybrids. New Phytol 184: 828–841. doi: 10.1111/j.1469-8137.2009.02964.x PMID: 19656303

17. Wang W, Xia H, Yang X, Xu T, Si HJ, Cai XX, et al. (2014) A novel 5-enolpyruvoylshikimate-3-phosphate (EPSP) synthase transgene for glyphosate resistance stimulates growth and fecundity in weedy rice (Oryza sativa) without herbicide. New Phytol 202: 679–688. doi: 10.1111/nph.12428 PMID: 23905647

18. Bartsch D, Lehnen M, Clegg J, Pohl-Orf M, Schuphan I, Ellstrand NC (1999) Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. Mol Ecol 8: 1733–1741. PMID: 10583835

19. van de Wiel C, van der Linden G, den Nijs H, Flavell A, Syed N, Jorgensen R, et al. (2003) An EU project on gene flow analysis between crop and wild forms of lettuce and chicory in the context of GMO biosafety: first results in lettuce. Proceedings of the Eucarpia meeting on leafy vegetables genetics and breeding, Noordwijkerhout, The Netherlands, 19–21 March, 2003. pp. 111–116.

20. Snow AA, Culley TM, Campbell LG, Sweeney PM, Hegde SG, Ellstrand NC (2010) Long-term persistence of crop alleles in weedy populations of wild radish (*Raphanus raphanistrum*). New Phytol 186: 537–548. doi: 10.1111/j.1469-8137.2009.03172.x PMID: 20122132

21. Davidar P, Snow AA, Rajkumar M, Pasquet R, Daunay M-C, Mutegi E (2015) The potential for crop to wild hybridization in eggplant (*Solanum melongena*; Solanaceae) in southern India. Am J Bot 102: 129–139. doi: 10.3732/ajb.1400404 PMID: 25587155

22. Gressel J (2015) Dealing with transgene flow of crop protection traits from crops to their relatives. Pest Manag Sci 71: 658–667. doi: 10.1002/ps.3850 PMID: 24977384

23. Committee on Genetically Engineered Crops (2016) Genetically Engineered Crops: Experiences and Prospects. The National Academies Press. 420 p. doi: 10.17226/23395

24. Heywood VH (1983) Relationships and evolution in the *Daucus carota* complex. Isr J Bot. 32: 51–65.

25. Hauser TP, Bjørn GK (2001) Hybrids between wild and cultivated carrots in Danish carrot fields. Genet Resour Crop Evol 48: 499–506.

26. Hauser TP (2002) Frost sensitivity of hybrids between wild and cultivated carrots. Conserv Genet 3: 73–76.

27. Iorizzo M, Senalik DA, Ellison SL, Grzebelus D, Cavagnaro PF, Allender C, et al. (2013) Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae) 1. Am J Bot 100: 930–938. doi: 10.3732/ajb.1300055 PMID: 23594914

28. Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, et al. (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet doi: 10.1038/ng.3565

29. Baranski R (2008) Genetic transformation of carrot (*Daucus carota*) and other Apiaceae species. Plant Breed. 2:18–38.

30. Umehara M, Eguchi I, Kaneko D, Ono M, Kamada H (2005) Evaluation of gene flow and its environmental effects in the field. Plant Biotechnol 22: 497–504.

31. Simon PW, Freeman RE, Vieira J V, Boiteux LS, Briard M, Nothnagel T, et al. (2008) Carrot: In:. Prohens M.J. Carena FN, editor. Handbook of Crop Breeding, Volume 1,Vegetable Breeding. Heidelberg, Germany: Springer-Verlag, GmBH. pp. 327–357.

32. Wijnheijmer EHM, Brandenburg WA, Ter Borg SJ (1989) Interactions between wild and cultivated carrots (Daucus carota L.) in the Netherlands. Euphytica 40: 147–154.

33. Harris S, Ingram R (1991) Chloroplast DNA and Biosystematics: The effects of intraspecific diversity and plastid transmission. Taxon 40: 393.

34. Reboud X, Zeyl C (1994) Organelle inheritance in plants. Heredity 72: 132–140.

35. Mogensen H (1996) The hows and whys of cytoplasmic inheritance in seed plants. Am J Bot 83: 383–404.

36. Röhr H, Kues U, Stahl U (1998) Organelle DNA of plants and fungi: inheritance and recombination. Prog Bot 60: 39–87.

37. Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14: 689–701. PMID: 15723661

38. Steinborn R, Linke B, Nothnagel T, Börner T (1995) Inheritance of chloroplast and mitochondrial DNA in alloplasmic forms of the genus Daucus. Theor Appl Genet 91: 632–638. doi: 10.1007/BF00223290 PMID: 24169891

39. Vivek BS, Ngo QA, Simon PW (1999) Evidence for maternal inheritance of the chloroplast genome in cultivated carrot (Daucus carota L. ssp. sativus). Theor Appl Genet 98: 669–672.

40. Mandel JR, McAssey EV, Roland KM, McCauley DE (2012) Mitochondrial Gene Diversity Associated with the atp9 Stop Codon in Natural Populations of Wild Carrot (Daucus carota ssp. carota). J Hered 103: 418–425. doi: 10.1093/jhered/esr142 PMID: 22337997

41. McCauley DE (2013) Tansley review: Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytol 200: 966–977. doi: 10.1111/nph.12431 PMID: 23952142

42. Breton S, Stewart DT (2015) Atypical mitochondrial inheritance patterns in eukaryotes. Génome / Cons Natl Rech Canada 58: 423–431.

43. Mandel JR, McCauley DE (2015) Pervasive mitochondrial sequence heteroplasmy in natural populations of wild carrot, Daucus carota spp. carota L. PLoS One 10: e0136303. doi: 10.1371/journal.pone.0136303 PMID: 26295342

44. Sherwin WB, Jabot F, Rush R, Rossetto M (2006) Measurement of biological information with applications from genes to landscapes. Mol Ecol 15: 2857–2869. doi: 10.1111/j.1365-294X.2006.02992.x PMID: 16911206

45. Sherwin WB (2015) Genes are information, so information theory is coming to the aid of evolutionary biology. Mol Ecol Resour 15: 1259–1261. doi: 10.1111/1755-0998.12458 PMID: 26452559

46. Smouse PE, Whitehead MR, Peakall R (2015) An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. Mol Ecol Resour 15: 1375–1384. doi: 10.1111/1755-0998.12422 PMID: 25916981

47. Campbell LG, Lee D, Shukla K, Waite TA, Bartsch D (2016) An ecological approach to measuring the evolutionary consequences of gene flow from crops to wild or weedy relatives. Appl Plant Sci 4: 1500114.

48. Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27: 623–656.

49. Rong J, Janson S, Umehara M, Ono M, Vrieling K (2010) Historical and contemporary gene dispersal in wild carrot (Daucus carota ssp. carota) populations. Ann Bot: 285–296. doi: 10.1093/aob/mcq108

50. Lacey E (1981) Seed dispersal in wild carrot (Daucus carota). Michigan Bot 20: 15–20.

51. Lacey E (1982) Timing of Seed Dispersal in Daucus Carota. Oikos 39: 83–91.

52. Hamilton M (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol Ecol 8: 521–523. PMID: 10199016

53. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18: 233–234. PMID: 10657137

54. Mandel JR, Milton EF, Donovan LA, Knapp SJ, Burke JM (2012) Genetic diversity and population structure in the rare Algodones sunflower (Helianthus niveus ssp. tephrodes). Conserv Genet 14:31–40.

55. Ellis JR, Bentley KE, McCauley DE (2008) Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower Helianthus verticillatus. Heredity 100: 574–580. doi: 10.1038/hdy.2008.11 PMID: 18301440

56. Cavagnaro PF, Chung S-M, Manin S, Yildiz M, Ali A, Alessandro MS, et al. (2011) Microsatellite isolation and marker development in carrot—genomic distribution, linkage mapping, genetic diversity

analysis and marker transferability across Apiaceae. BMC Genomics 12: 386. doi: 10.1186/1471-2164-12-386 PMID: 21806822

57. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539. PMID: 22820204

58. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590. PMID: 17248844

59. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491. PMID: 1644282

60. Wright S (1951) The genetical structure of populations. Ann Eugen 15: 323–354. PMID: 24540312

61. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959. PMID: 10835412

62. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611–2620. PMID: 15969739

63. Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4: 359–361.

64. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806. PMID: 17485429

65. Rosenberg NA (2003) distruct: a program for the graphical display of population structure. Mol Ecol Notes 4: 137–138.

66. Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177–1191. PMID: 12663554

67. Ronfort J, Saumitou-Laprade P, Cuguen J, Couvet D (1995) Mitochondrial DNA diversity and male sterility in natural populations of Daucus carota ssp carota. Theor Appl Genet 91: 150–159. doi: 10.1007/BF00220872 PMID: 24169681

68. Wade M, Mccauley D (1988) Extinction and Recolonization: Their Effects on the Genetic Differentiation of Local Populations on JSTOR. Evolution (N Y): 995–1005.

69. Whitlock M, Mccauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution (N Y) 44: 1717–1724.

70. Boblenz K, Nothnagel T, Metzlaff M (1990) Paternal inheritance of plastids in the genus Daucus. MGG Mol Gen Genet 220: 489–491.

71. Szklarczyk M, Szymański M, Wójcik-Jagła M, Simon PW, Weihe A, Börner T (2014) Mitochondrial atp9 genes from petaloid male-sterile and male-fertile carrots differ in their status of heteroplasmy, recombination involvement, post-transcriptional processing as well as accumulation of RNA and protein product. Theor Appl Genet 127: 1689–1701. doi: 10.1007/s00122-014-2331-x PMID: 24913361

72. Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. Trends Ecol Evol 18: 411–417.

73. Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol 168: 39–50. PMID: 16159319

74. Xu J (2005) The inheritance of organelle genes and genomes: patterns and mechanisms. Génome / Cons Natl Rech Canada 48: 951–958.

75. Burt A, Trivers R (2006) Genes in conflict: the biology of selfish genetic elements. Cambridge, Massachusetts: Belknap Press of Harvard University Press.