Genetic Diversity Among Mitochondrial Genes of *Daucus carota*

Katherine M. Roland, Edward V. McAssey, and David E. McCauley

KEYWORDS. Mitochondrial DNA, genetic diversity, *Daucus carota*, population structure.

BRIEF. The quantitative genetic diversity of select mitochondrial genes and population structure of *Daucus carota* populations was studied.

ABSTRACT. The genetic diversity of the plant mitochondrial genome has not been well studied in natural populations, as it was assumed to lack diversity due to strictly maternal inheritance. However, previous studies have shown diversity in specific genes in the mitochondrial DNA of Daucus carota, including both size differences and single-nucleotide polymorphisms (SNP). To continue the study of mitochondrial genetic diversity, five-locus polymerase chain reaction/restriction fragment length polymorphism (PCR/RFLP) genotypes (based on the mitochondrial genes atp1, atp9, nad, and cob) were determined for 35 individuals collected from six North American populations. Six five-locus PCR/RFLP genotypes were found. The frequency of each genotype was determined within each population, and within the total data set. Overall, the genotypic frequencies within the total data set were roughly equal. An average of two genotypes were found per population. Intergenic recombination has the potential to create novel multilocus genotypes in plant mitochondrial genomes, and may account for the aforementioned genetic diversity. This may indicate biparental inheritance, rather than strictly maternal inheritance, which is a departure from the common explanation of mitochondrial inheritance.

INTRODUCTION.

Biodiversity plays an important role in the natural environment, but also in agriculture, human health, and even business. It is how an ecosystem stays stable, how mass extinctions are avoided, and how new species are able to evolve. An important part of biodiversity is genetic diversity, which is responsible for most evolution. Understanding how genetic diversity is achieved is vital for understanding both evolution and how biodiversity is maintained. Genetic diversity is found in both the nuclear genome and the cytonuclear genomes, the DNA in the mitochondria and chloroplast.

The plant mitochondrial genome has been of increasing interest to evolutionary biologists for a variety of reasons. Mitochondrial genetic markers have been used in studies of population structure and gene flow by seeds [1], as markers in phylogenetic studies [2][3], and mitochondrial genetic mutations are important in sex determination of many gynodioecious species [3].

For many years, the mitochondrion was thought to be strictly maternally inherited in plants, and that there was little to no diversity in the mitochondrial genome. Recent studies have shown that there can be significantly more diversity than was predicted [4]. In fact, recent studies have shown that the species Silene vulgaris is at least occasionally subject to paternal leakage, when some or all of the mitochondrial genome is inherited from the pollen donor, which resulted in widespread heteroplasmy [4,5]. Heteroplasmy is a phenomenon in which there are two distinct copies of an organellar DNA within a single cell. Heteroplasmy can induce recombination, when one copy of a genome incorporates pieces of the other. Since the mitochondrial genome was thought to be maternally inherited, there would only be one distinct genome within every cell, and recombination, the cause of most genetic diversity, would not occur.

The idea that the mitochondrial genome could have more diversity than is assumed could greatly impact several areas of research. Markers in the mitochondrial genome are commonly used in population structure and gene flow studies, which are generally based on statistics that assume a lack of diversity and strict maternal inheritance. Consequentially, many previous studies may have come to inaccurate conclusions. More important in the evolutionary aspect, genetic diversity in the mitochondrial genome would have an impact on gynodioecious plants that have been found to have cytonuclear male sterilization (CMS) genes in the mitochondrial genome. CMS elements are genes located in the mitochondrial genome that sterilize the male components in hermaphroditic plants. They are of interest because CMS elements are an example of genetic conflict, between the CMS genes and restorer genes in the nuclear genome [3]. Restorer genes counteract the effects of the CMS elements to restore male function.

The species Daucus carota (Figure 1) is a weedy exotic, originating in northern Europe, and is a gynodioecious plant that has a male sterile form. It is very closely related to the domesticated carrot and is found abundantly across North America. It is of interest because of its close relation to the domesticated carrot, that it is gynodioecious, and that it has been shown to have some diversity in the mitochondrial genome. Previous studies have shown diversity in four specific genes, in five loci, in the mitochondrial DNA, which include both size differences and single-nucleotide polymorphisms (SNPs) [6]. The diversity at these loci was common enough to reliably find variation within several populations. The previous study of these loci makes it likely that there will be some limited diversity both within each population and between different populations. The goal of this research is to quantify genetic diversity in five previously unstudied D. carota populations. This will expand the understanding of mitochondrial diversity in plants and show whether there is evidence for recombination or paternal leakage in D. carota. Individuals from five distinct populations will be genotyped for the five genetic polymorphisms, and then tests for population structure will be applied. Previous experiments indicate that there will be genetic diversity, though do not demonstrate the extent. It is most likely that there is still somewhat strict population structure.

MATERIALS AND METHODS.

Study System.

D. carota is a weedy exotic species native to Europe, and introduced to North America. It is biennial, can grow up to one meter tall, and flowers from June to August. Variation has been identified in six loci in the mitochondrial DNA (mtDNA) of D. carota.

Collection Method.

Between two and eight individual samples were collected according to two methods from six unique populations. Leaf material was either collected from individuals grown to seedlings in a greenhouse, or leaf material was collected from plants in the wild.

Seeds were collected from individual plants and reared into seedlings in a greenhouse. Approximately 100g of leaf material was collected for DNA extraction from each individual. This method was applied in three populations, all from different areas in New York.

In the second method, one leaflet was collected from each individual. These were stored in a 13°C refrigerator for up to one week, and then approximately 100g leaf material was collected for DNA extraction. This method was applied in three populations, two from Tennessee and one from Georgia

Genetic Material.

Extractions were conduced using a variation of the DNeasy Plant Mini Kit (Qiagen) in 100□L reactions. Cell lysis was performed using sand and a pestle, rather than with liquid nitrogen. Emulsion was performed with 100□L water.

PCR

A 50.□L PCR reaction was performed for each individual at each locus. PCR was performed using the method and primers described in [6].

RFLP.

A 20□L RFLP digestion was performed using a specific enzyme for each individual at each polymorphism, excluding atp1, for which it was unnecessary to digest. Digestions were incubated for at least one hour at specified temperatures. Results were imaged on 2% Ultrapure Agarose gels stained with ethidium bromide for five to ten minutes (including atp1 PCR product). Gels were run at 70 volts for one to two hours.

Genotyping.

atp1

A PCR reaction was performed to amplify the non-coding region just outside the atp1 gene and to determine the presence of one, two, or three copies of a repeat (denoted N1, Y1, and Y1Y1, respectively. Size difference was determined by imaging the PCR product directly.

nad

A single-nucleotide polymorphism (SNP) exists just outside the gene nad6. The New England Biolab (NEB) restriction enzyme ApoI was combined with NEB buffer 3 with BSA. This digestion reaction was run at 50°C. This would either cut or leave uncut the DNA fragment containing the SNP. The polymorphism was described as either cut or uncut (U and C, respectively).

cob

A SNP occurs outside the gene cob. NEB restriction enzyme BfaI was combined with NEB buffer 4. This digestion was run at 37°C. It would either cut the DNA fragment into four (uncut, or U) or five (cut, or C) bands.

atp9

This gene contains both a repeat difference in the adjacent non-coding region and a SNP near the end of the gene. The repeat difference was determined using NEB restriction enzyme MspI and NEB buffer 4, run at 37°C. The difference of fragment size, due to one copy or two copies of the repeat, was denoted as N9 and Y9, respectively.

The SNP difference was determined using NEB enzymes DraI and ApoI. These two enzymes were used in separate reactions. DraI was combined with NEB buffer 4 and run at 37°F. This enzyme cut the DNA fragment if the full codon was TAA. ApoI was combined with NEB buffer 3 and run at 50°C with BSA. This enzyme cut the DNA fragment if the full codon was either TAA or CAA. Considered together, these two digestions made it possible to classify the codon as TAA (T), CAA (C), or AAA (A).

Genotype Construction.

Once all five specific locus genotypes for each individual were known, five-locus genotypes were constructed. These overall genotypes were composed in the following order: atp1, atp9 repeat, atp9 SNP, nad, and cob. Individual loci were denoted as described previously. For example, the genotype "N1N9TCU" would indicate an individual with one copy of the repeat in atp1 (N1), one copy of the repeat in atp9 (N9), a TAA codon in the atp9 SNP (T), a nad digestion that cut (C), and a cob digestion that had five, rather than seven, DNA bands (U).

Analysis of Population Structure.

Relative frequencies in percent of the whole of each unique genotype were constructed for the total data set by dividing the number of individuals with a certain genotype by the total number of individuals. This process was repeated for unique genotypes within each population.

The fixation index (FST) was calculated between the most common five-locus genotype (N1N9CUC) and the other genotypes combined to calculate population structure (McCauley and Ellis 2008). It yields a number between one and zero, a value of one implying complete separation of populations and a value of zero implying no population structure. FST was calculated according to the following three equations,

$$p_{bar} = \sum(p_i) / a, \quad (1)$$

$$\sigma p_i = \sum(p_i - p_{bar})^2 / (a-1), \quad (2)$$

$$FST = \sigma p_i / (p_{bar} * [1-p_{bar}]), \quad (3)$$

where p_i equals the frequency of the genotype N1N9CUC for each population, p_{bar} equals the average frequency of the genotype across all populations, and a is the number of populations.

The four gamete test for recombination [8] was also performed between each pair of genes. This is done by cataloguing the number of combinations of alleles of any pair of genes that is actually present in the data set. Recombination must occur for all possibilities to be present.

RESULTS.

Sample Collection.

Samples were collected from six unique populations: Bowl, Church, Grand Gorge, Shelby, UT, and White Shoals. In the population Bowl, seven individuals were collected. In Church, eight individuals were collected. Three individuals were collected from Grand Gorge. Eight individuals were collected from Shelby. Seven individuals were collected from UT, and two individuals were collected from White Shoals.

Genetic Markers.

The data can be found in the supplemental material. Examples of a repeat difference and SNP gel are depicted in Figure 1.

Genotype Construction.

Given the genetic variation found within this set of genes, there are 72 possible five-locus genotypes, of which six were found. The following are the labels of the genotypes found: N1N9CUC, Y1Y1Y9TUU, Y1N9CUC, N1N9ACC, Y1N9ACC, and Y1Y9TUU. Within the combined data set, the genotypic frequency of each of the five-locus genotypes was roughly equal to 16.7%, or 1/6 of the total (Figure 1). Each genotype was represented by at least four individuals. An average of two and no more than three genotypes was found in a single population (Figure 2).

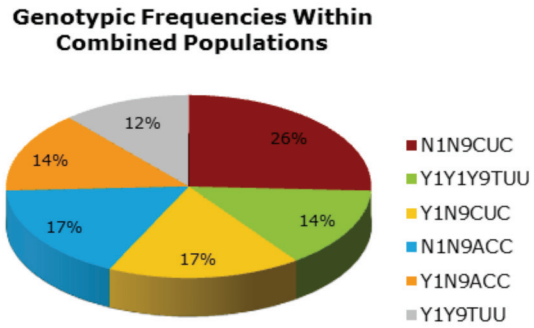

Figure 1. Frequency of each five-locus genotype within the combined data set, including all six populations and a total of 35 individual *Daucus carota* plants. The genotypic frequencies were approximately 16.7% of the total population. These results show that there exists genetic diversity of the mitochondrial genes.

Analysis of Population Structure.

Population structure was calculated between all six populations. The FST was calculated for the most frequent genotype (N1N9CUC) set against a combination of the other genotypes. This was evaluated to be 0.43, indicating moderate population structure (FST=0.43). The four-gamete test, performed between each pair of genes, showed no evidence for recombination. It did show a strict association between the atp9 SNP and nad, and the atp9 SNP and cob (TU, CU, AC, and TU, CC, AC, respectively).

DISCUSSION.

While other studies have shown that the mitochondrial genome is much more diverse than originally predicted [7], and that recombination does occur indicating paternal leakage, this study suggests that there is some unexpected diversity but not to the extent shown in other species. The FST was found to be 0.43, indicating moderate population structure. This number agrees with the other findings of approximately two genotypes per population, which indicates mild diversity within populations. This degree of structure is lower than is generally predicted for the mitochondrial genome, but the FST is still much higher than the average value for nuclear genes.

The number of individuals collected from each population was variable. This accounts for the nearly equal number of individuals found with each genotype. The four-gamete test showed no evidence for recombination, indicating that there was no paternal leakage. This does not mean that it does not occur, because while a small sample size was enough to do a small survey of genetic frequency, rare cases of paternal leakage would be difficult to detect. Several cases of heteroplasmy, excluded from this study, suggest the existence of some paternal leakage. A larger sample size, with more populations and more individuals per population, would be a better test for paternal leakage, biparental inheritance, and recombination.

The genetic diversity of mitochondrial genes in D. carota were shown to be diverse and less structured than would be expected when strict maternal inheritance occurs. This study contributed to the understanding of the mitochondrial genome, and genetic diversity in general. While it is only on one specific species, it shows that the genetic diversity among mitochondrial genes is variable according to the species in question.

Any research on the plant mitochondrion is useful simply because of the general lack of such research. Knowledge of the degree of diversity in the mitochondrial DNA is necessary, partially because of its use as a marker of seed dispersal and gene flow. Also, the close relationship of D. carota to the domesticated carrot means that any research done on its wild relative is potentially beneficial to the agricultural industry.

If more time were available, an expanded version of this study could further illuminate the diversity among mitochondrial genes in D. carota. The limited sample size of this study could have made it difficult to find uncommon combinations of genotypes that could have completed the four gamete test. It would also benefit the validity of this study to sequence several representatives of each genotype for each marker to ensure that the PCR/RFLP data is accurate, and that there are not even more genotypes for a particular marker that are similar in size once genotyped. Also, a modification of these methods could also be used on different species to further expand knowledge of the mitochondrial genome.

More studies could be performed that use the four gamete test to compare mitochondrial genes with chloroplast genes, as both were thought to be maternally inherited and both have shown evidence of paternal leakage and increased diversity. Data from this study could be combined with new data from the same individuals for chloroplast gene markers to achieve this.

Biological diversity is important because it enables evolution and adaptation to change, which are vital for ecological survival. Genetic diversity is a key component to biological diversity. This study contributes to the general understanding of genetic diversity, specifically within the mitochondrial genome. This study corroborates prior studies indicating that the previous understanding of mitochondrial diversity was incomplete, and that mitochondrial diversity is a factor that must be considered to properly understand genetic diversity in plant species.

ACKNOWLEDGMENTS. I would like to thank Jennifer Ellis for providing information about the primers and enzymes involved. I would also like to thank Jonathan Ertlet for his generous assistance in the Vanderbilt University greenhouse, Steven J. Baskauf for the donation of the Daucus carota photograph, and all the instructors at the School for Science and Math at Vanderbilt.

SUPPORTING INFORMATION.

Extended PCR and RFLP methods

Figure S1. Picture of *Daucus carota*.

Figure S2. Complete genotypic data set for all individuals.

Figure S3. Genotypic frequencies within each population.

REFERENCES.

1. R. J. Petit, J. Duminil, S. Fineschi, A. Hampe, D. Salvini, et al, Molec. Ecol. 14, 689 (2005).
2. T.J. Barkman, *et al., PNAS*. **97**, 13166 (2000)
3. D. E. McCauley, M. F. Bailey, An. of Bot. **104**, 611 (2009).
4. D. E. McCauley, M. F. Bailey, N. A. Sherman, M. Z. Darnell, Heredity. **95**, 50 (2005).
5. M. E. Welch, M. Z. Darnell, D. E. McCauley, Genetics. **174**, 829 (2006).
6. I. C. Bach, A. Olesen, P. W. Simon, Euphytica. **127**, 353 (2002).
7. D.E. McCauley, J.R. Ellis, *Evolution*. **62**, 823(2008).
8. R. R. Hudson, N. L. Kaplan, Genetics. **111**, 147 (1985).

Katherine Roland is a student at Hume-Fogg Academic Magnet High School and enrolled in the School for Science and Math at Vanderbilt