ORIGINAL ARTICLE

Patterns of cyto-nuclear linkage disequilibrium in *Silene latifolia*: genomic heterogeneity and temporal stability

PD Fields¹, DE McCauley², EV McAssey^{2,3} and DR Taylor¹

Non-random association of alleles in the nucleus and cytoplasmic organelles, or cyto-nuclear linkage disequilibrium (LD), is both an important component of a number of evolutionary processes and a statistical indicator of others. The evolutionary significance of cyto-nuclear LD will depend on both its magnitude and how stable those associations are through time. Here, we use a longitudinal population genetic data set to explore the magnitude and temporal dynamics of cyto-nuclear disequilibria through time. We genotyped 135 and 170 individuals from 16 and 17 patches of the plant species *Silene latifolia* in Southwestern VA, sampled in 1993 and 2008, respectively. Individuals were genotyped at 14 highly polymorphic microsatellite markers and a single-nucleotide polymorphism (SNP) in the mitochondrial gene, *atp1*. Normalized LD (D') between nuclear and cytoplasmic loci varied considerably depending on which nuclear locus was considered (ranging from 0.005–0.632). Four of the 14 cyto-nuclear associations showed a statistically significant shift over approximately seven generations. However, the overall magnitude of this disequilibrium was largely stable over time. The observed origin and stability of cyto-nuclear LD is most likely caused by the slow admixture between anciently diverged lineages within the species' newly invaded range, and the local spatial structure and metapopulation dynamics that are known to structure genetic variation in this system. *Heredity* (2014) **112**, 99–104; doi:10.1038/hdy.2013.79; published online 4 September 2013

Keywords: cyto-nuclear; linkage disequilibrium; metapopulations; spatiotemporal; *Silene*

INTRODUCTION

Population genetic studies generally focus on either the nuclear or the cytoplasmic (mitochondria or chloroplast) genome. Additional insights can be gained by measuring joint allelic associations at loci derived from nuclear and cytoplasmic genomes, or cyto-nuclear linkage disequilibrium (LD). Analyses of cyto-nuclear LD can document evolutionary and demographic trends in population genetics. Processes like structure (Nei and Li, 1973), gene flow (Asmussen and Schnabel, 1991) and non-random mating/inbreeding (Asmussen *et al.*, 1989; Slatkin, 2008) all influence cyto-nuclear LD. Selection-based treatments of cyto-nuclear LD focus on the importance of epistatic interactions between the nuclear and cytoplasmic genomes (Wade and Goodnight, 2006; Brandvain and Wade, 2009). Theoretical investigations of genetic drift predict an influence of population size on the variance of LD measurements (Datta *et al.*, 1996). Additionally, demographic processes have been shown to greatly facilitate non-random associations between the nucleus and a cytoplasmic genome. For example, cyto-nuclear LD (or lack thereof) is an informative metric when considering admixture of two populations of the same species (Arnold, 2006). Thus, cyto-nuclear LD will be driven by the interaction of spatiotemporal selective and demographic parameters specific to a system of study, or by non-equilibrium population dynamics associated with the metapopulation structure of many natural systems.

While the aforementioned investigations provide substantial information concerning the patterns resulting from cyto-nuclear disequilibrium, few incorporate a temporal component. A pioneering study in

Drosophila montana used allozyme polymorphisms to track changes in nuclear LD over a 5-year time period (Baker, 1975). The author concluded that maintenance of LD was consistent with epistatic selection maintaining favorable gene–gene interactions (Baker, 1975). Conversely, in *Drosophila melanogaster* it was found that there was an absence of LD among allozyme loci when samples were taken at two time points within the same year (Langley *et al.*, 1977). More recent studies in *Caenorhabditis elegans* have suggested that the maintenance of significant LD among nuclear-based microsatellite markers over short timescales could be generated by genetic drift in small populations or selection against hybrid progeny (outbreeding depression) (Barrière and Félix, 2007).

While the previously described animal-oriented studies focused on LD among nuclear markers (allozymes and microsatellites), the development of markers in cytoplasmic genomes has allowed for an in-depth analysis of cyto-nuclear LD. Thus, tracking levels of cyto-nuclear LD on a microevolutionary timescale allow for an analysis of the tempo of the change in multi-locus interactions between the nuclear and organellar genomes. Of particular interest is determining whether or not a cyto-nuclear interaction is deteriorating, or strengthening over time. Deterioration of LD can be achieved with sexual recombination although it should be noted that population structure (bi-parental inbreeding) can slow down the rate of deterioration due to outcrossing (Brandvain and Wade, 2009). Strengthening of LD can be the result of selection for a particular set of alleles (epistasis), demographic expansion where population sizes grow rapidly or due to a genetic bottleneck where only a small

¹Department of Biology, University of Virginia, Charlottesville, VA, USA and ²Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

³Current Address: Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

Correspondence: PD Fields, Department of Biology, University of Virginia, PO Box 400328, Charlottesville, VA 22904, USA.

E-mail: pdf8z@virginia.edu

Received 30 September 2012; revised 21 June 2013; accepted 1 July 2013; published online 4 September 2013

number of allelic associations remain after a stochastic event and thus particular multi-locus genotypes rise in number despite an absence of selection. It is essential that demographic and selective forces be separated when trying to infer the evolutionary dynamics of a particular system (Stinchcombe and Hoekstra, 2007). As demographic processes generally act on the whole genome, locus-specific patterns of maintenance or strengthening of LD are suggestive of selection for a particular cyto-nuclear combination of alleles.

Silene latifolia is a dioecious, insect-pollinated plant that has been the subject of a considerable number of studies that focus on local population genetic structure and the effects of metapopulation dynamics on that structure (McCauley, 1994; McCauley *et al.*, 1995, 1996; Bernasconi *et al.*, 2009). Many of these studies were conducted >15 years ago, before the introduction of most of the DNA-based genetic markers used in studies of population structure today. However, they did utilize PCR-based chloroplast DNA (cpDNA) markers for comparison with allozyme nuclear markers (the workhorse of empirical population genetics at that time). McCauley (1994) observed a contrast between the moderate local population structure of the nuclear gene markers ($F_{ST} \sim 0.20$) and a much greater degree of structure of the chloroplast markers ($F_{ST} \sim 0.60$). He attributed this to differences in the mode of dispersal of the two genomes (seed + pollen movement in the bi-parentally inherited nuclear genome versus seed movement in the maternally inherited chloroplast genome) (McCauley, 1994). Further, founding events associated with the repeated establishment of new populations enhanced the magnitude of structure for both genomes (McCauley *et al.*, 1995). These metapopulation dynamics have since been shown to be important determinants of population genetic structure in several other systems (Giles and Goudet, 1997).

The earlier studies of *S. latifolia* did not focus on cyto-nuclear LD, but the DNA extractions were archived and can now be used to generate nuclear and cyto-nuclear genotypes using the more powerful DNA-based genotyping techniques that are currently available. A comparison of cyto-nuclear associations between archived samples and current *S. latifolia* collections allow us to examine the temporal dynamics of cyto-nuclear LD in a natural metapopulation. Standard questions about the magnitude of cyto-nuclear LD, and locus-to-locus variation in LD, can be extended to include their temporal dynamics. Here, we report on the association between variants found in a cytoplasmic (mitochondrial) SNP marker and variants found in 14 nuclear genes in samples of natural populations of *S. latifolia* taken in 1993 and again in 2008, including observations of locus-to-locus heterogeneity of cyto-nuclear LD and the temporal stability of those associations.

We also include a full population genetic analysis of our nuclear markers. By doing so, we allow for a larger comparison to many other published studies of nuclear–nuclear LD, suggesting that the presence and stability of cyto-nuclear LD may be far more common than has previously been suggested.

MATERIALS AND METHODS

Collections of *S. latifolia* were first made from 16 patches of plants (with a mean, or $\bar{N}$, per site of 8.5, and a range of 1–29 individuals per patch) found along the roadsides of Giles County, VA, in summer 1993. All patches were located <20 km from one another. These samples represent a subset of the individuals included in earlier studies of local population structure (McCauley, 1994; McCauley *et al.*, 1995). For those studies leaf tissue was used both to generate nuclear allozyme genotypes and cpDNA RFLPs. Genomic DNA used for cpDNA genotyping was extracted using standard methods. This DNA was then stored at -80°C until it was genotyped recently for the mtDNA and microsatellite markers used in the current study.

The same population networks were sampled (17 in total, and $\bar{N} = 10$ per patch) in the summer of 2008. *S. latifolia* has been described as a short-lived perennial (Bernasconi *et al.*, 2009) and it has been our experience from field experiments that the average lifespan of *S. latifolia* is ~ 2 years with the age at first reproduction being considerably less. Thus, we estimate that a minimum of seven generations separates the collections, and are confident that no individual plants were sampled twice. It should be noted that while the two collections were made from the same stretch of the metapopulation, they were not necessarily made from exactly the same places. In fact, because this weedy plant is known to undergo fine-scale episodes of extinction and colonization (McCauley *et al.*, 1995), some local populations collected in 1993 may have become extinct by 2008. Similarly, some 2008 collections may have been from recent colonizations of localities that did not contain *S. latifolia* in 1993.

For the 2008 collections, DNA was extracted from leaf tissue using the method described by Keller *et al.* (2012). All individuals from both collections were assayed for a SNP known to occur in the mitochondrial gene *atp1* by a PCR/RFLP method (McCauley and Ellis, 2008). We prefer this mitochondrial SNP as our cytoplasmic marker, rather than the cpDNA cytoplasmic markers used in earlier studies (for example, McCauley, 1994), because the cpDNA markers consisted of indel polymorphisms that could be more subject to homoplasy. The SNP in question determines the presence or absence of an *AluI* restriction enzyme cut site. Genomic DNA was subject to a PCR that utilized the *atp1* primers and cycle conditions used by McCauley and Ellis (2008). The resulting PCR product (10 μl) was digested with *AluI* using the manufacturers recommended conditions (NEB). The resulting fragments were electrophoresed on a 4% Metaphor agarose gel (Lonza Inc, Rockland, ME, USA), which was then stained with ethidium bromide for visualization.

Genomic DNA from the same individuals was used to generate multi-locus genotypes at 14 unlinked microsatellite loci. Microsatellites were derived from multiple sources (Moccia *et al.*, 2009; Abdoulaye *et al.*, 2010). PCR amplification was conducted using published methods for each marker. PCR products were amplified with the forward primer end-labeled with a fluorescent dye, either 5(or 6)-FAM, NED, TAMRA, JOE or VIC. Three to four PCR products of different loci were then pooled together and added to a loading buffer containing formamide and GENESCAN 400HD ROX size standard (Applied Biosystems). Following 5 minutes of denaturing at 95°C , fluorescently labeled fragments were separated on an Applied Biosystems 3130 sequencer and analyzed with GENEMAPPER v3.0 software (Applied Biosystems, Grand Island, NY, USA). Alleles were binned using the software TANDEM (Matschiner and Salzburger, 2009).

We calculated the observed (H_O) and expected (H_E) heterozygosity of our nuclear genetic markers using the software GenoDive version 2.0b21 (Meirmans and Van Tienderen, 2004). Estimates of genetic substructure using hierarchical F -statistics were calculated using the software FSTAT 2.9.3.2 (Goudet, 2002), with significant deviations from panmixia assessed by testing for Hardy–Weinberg Equilibrium with 10 000 permutations and $\alpha = 0.05$.

To estimate patterns of nuclear–nuclear LD (hereafter, nuclear LD) among microsatellite loci, we applied Hedrick's (1987) multi-allelic extension of Lewontin's (1964) normalized D' . This measure of LD is preferable over others as it is widely used, and given its normalization, can be compared with other studies (Slate and Pemberton, 2007; Li and Merila, 2010). D' ranges from zero (no allelic associations between loci) to one (complete allelic associations at two loci), though estimation of the statistic may be sensitive to allele frequencies and sample sizes (Slate and Pemberton, 2007; Li and Merila, 2010; but see Zapata, 2011). D' between two multi-allelic markers was calculated following Li and Merila (2010) equations (1) and (3), using Multiallelic interallelic disequilibrium analysis software (MIDAS) (Gaunt *et al.*, 2006).

Statistical significance of nuclear LD between pairs of loci for a given sampling date was estimated, under the null hypothesis of random allelic assortment, using a Monte-Carlo approximation of Fisher's exact test implemented in the software Arlequin (Excoffier and Lischer, 2010). Arlequin uses a Markov chain extension of Fisher's exact test for R_{xC} contingency tables (Slatkin, 1994; Li and Merila, 2010). 100 000 alternative tables were explored by the Markov chain (Slate and Pemberton, 2007; Li and Merila, 2010).

Cyto-nuclear LD was estimated between each nuclear microsatellite locus and the *atp1* mtDNA locus. Analysis followed the approach of Basten and Asmussen (1997), using the program CNM to estimate a standardized

estimate an allelic D' between each nuclear locus and *atp1* mtDNA locus. CNDm uses a Monte-Carlo approach to approximate Fisher's exact test for R_{xC} contingency tables and tests for significant deviations from the null hypothesis of no allelic association (Basten and Asmussen, 1997). For this analysis, all nuclear loci were treated as bi-allelic by pooling all alleles other than the most common allele in 1993 into a single composite allele (Latta *et al.*, 2001). This approach is preferable for the present analysis, as it provides a single value for each locus–locus pair, while also generating the most intermediate allele frequencies, thereby maximizing the bounds on D' . Additionally, given finite sample sizes, and the propensity of microsatellite mutation rates to generate a large class of rare, private alleles, which will rarely be in linkage equilibrium, our binning procedure preserves statistical robustness.

Cyto-nuclear LD was calculated separately for the 1993 and 2008 collection. We also calculated cyto-nuclear LD of a pooled sample of both years, which allows one to detect long-term patterns of cyto-nuclear LD and its consequences. For example, year-to-year consistency of individual D' values will reinforce one another, generating increased significance of cyto-nuclear LD. Conversely, when D' reverses between years, the overall significance of cyto-nuclear LD would be canceled out in a single pooled value. Year-to-year stability of D' values was evaluated statistically by two methods. Temporal consistency in the relative rankings of locus-specific D' values was evaluated by estimating the between-year correlation of the 14 pairs of locus-specific D' values. Recall that each locus-specific D' was based on observations of four possible cyto-nuclear genotype combinations (two mitochondrial variants combined with the common or binned minor nuclear alleles). This yields a 2×2 table of D' values in which each of the four entries has the same absolute value (two positive and two negative). The CNDm program summarizes this by reporting an absolute D' value for each year, and pooled across years (Table 2). These absolute values are not suitable, however, for testing the consistency of year-to-year associations, as such comparisons would not be sensitive to changes in the sign of an allele-specific value of D' between years. Thus, when calculating the correlation mentioned above we consider for each locus the sign of the D' value specific to the association of the most common nuclear allele with the most common mitochondrial allele in 1993 and that same combination in 2008.

Year-to-year heterogeneity was also tested for statistical significance on a nuclear locus-specific basis by noting that D' is mathematically similar to a product–moment correlation coefficient in that both consist of a covariance standardized to range from -1 to 1 . We used the Z-transformation approach suggested by Sokal and Rohlf (2012) for testing for heterogeneity among pairs

of correlation coefficients (that is, 1993 versus 2008 for each of the 14 loci). Heterogeneity among the 14 pooled D' values was also tested using this method. The pairwise test employs a t -statistic while the test across all 14 values employs a χ^2 statistic (Sokal and Rohlf, 2012).

The significance of multiple comparisons of cyto-nuclear LD among markers within years, and across years, was assessed using a Benjamini–Hochberg correction for false discovery (Benjamini and Hochberg, 1995). This approach has been suggested as a powerful analog Bonferroni correction, and has recently been used for nuclear–nuclear LD comparisons (Andras and Ebert, 2013).

RESULTS

Cyto-nuclear genotypes were obtained for 305 individuals (135 from the 1993 collection and 170 from the 2008 collection). For the *atp1* SNP, the common variant occurred at a frequency of 0.53 in 1993 and 0.58 in 2008. Table 1 presents a summary of global estimates of nuclear genetic diversity, population substructure (F_{IT}) and a summary of nuclear LD for the 1993 and 2008 samples. Individual microsatellite markers were consistently polymorphic between years, with an overall range of 3–22 alleles per locus. Additionally, individual loci show a large range of H_{O} (0.11–0.84) and expected H_{E} (0.08–0.90). Average $F_{\text{IT},1993} = 0.353$ (range -0.041 to 0.780) and $F_{\text{IT},2008} = 0.322$ (range 0.101 – 0.792), with 13 and 14 markers, respectively, exhibiting statistically significant ($P \leq 0.05$) deviations from panmixia, corroborating deviations observed in H_{E} . In the 1993 and 2008 samples, there was an average of eight statistically significant non-random associations between each nuclear marker locus and the other 13 nuclear loci. Specific D' values and statistical significance of the deviation from random association for each marker pair is described in Supplementary Tables 1 and 2, for 1993 and 2008, respectively.

Table 2 presents the 1993, 2008 and pooled year absolute D' values between each of the 14 nuclear genes and the mitochondrial *atp1* SNP. Inspection of Table 2 shows that there was little change in the overall levels of cyto-nuclear LD between 1993 and 2008. The absolute magnitude of D' increased for eight loci and decreased for six (average locus-specific $|D'|$ in 1993 = 0.266 , average $|D'|$ in 2008 = 0.263). The mean absolute pooled D' value was 0.230 . Locus-specific pooled

Table 1 Global estimates of genetic diversity and population substructure within temporally separated samples

Locus	1993							2008						
	M	N	P	H _O	H _E	F _{IT}	LD	N	P	H _O	H _E	F _{IT}	LD	
SL_eSSR04	3	8	0.43	0.58	0.74	0.244***	7	8	0.55	0.57	0.65	0.143***	7	
SL_eSSR06	3	15	0.32	0.50	0.82	0.405***	9	17	0.22	0.71	0.84	0.178***	10	
SL_eSSR09	3	8	0.74	0.32	0.44	0.282***	8	8	0.75	0.39	0.43	0.101*	7	
SL_eSSR12	2	12	0.22	0.72	0.87	0.185***	10	13	0.39	0.69	0.80	0.142***	11	
SL_eSSR16	3	9	0.54	0.40	0.63	0.401***	10	6	0.54	0.48	0.62	0.245***	9	
SL_eSSR20	3	5	0.94	0.12	0.11	−0.041	1	5	0.96	0.07	0.08	0.140*	1	
SL_eSSR29	3	17	0.19	0.74	0.89	0.180***	11	22	0.22	0.84	0.90	0.078***	8	
SL_eSSR30	3	17	0.43	0.69	0.77	0.111**	9	18	0.36	0.43	0.81	0.489***	11	
slat_18	3	9	0.36	0.39	0.76	0.498***	10	11	0.43	0.37	0.76	0.519***	7	
slat_32	4	7	0.34	0.44	0.78	0.455***	11	8	0.30	0.32	0.78	0.608***	11	
slat_33	3	3	0.55	0.13	0.50	0.780***	12	2	0.65	0.11	0.46	0.792***	13	
slat_48	3	5	0.45	0.27	0.66	0.597***	3	6	0.69	0.17	0.46	0.662***	7	
slat_72	3	11	0.46	0.48	0.74	0.359***	9	13	0.19	0.38	0.80	0.582***	12	
slat_85	3	17	0.44	0.54	0.76	0.303***	12	10	0.58	0.29	0.60	0.528***	8	
Overall	—	10.21	0.46	0.45	0.68	0.353***	8.71	10.5	0.49	0.42	0.64	0.377***	8.71	

M = repeat size of the marker; N = number of observed alleles; P = allele frequency of the most frequent allele.

H_{O} and H_{E} are observed and expected heterozygosity, respectively, of each sampled microsatellite locus; F_{IT} is the correlation of alleles within individuals relative to the metapopulation as a whole, and can be interpreted as an approximate measure of deviation from panmixia. Statistical deviations from panmixia are indicated by the symbols * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ and LD is an indication of the number of other nuclear loci for which each focal nuclear locus exhibits a non-random association, based upon an approximation of Fisher's exact test of random association and a $P \leq 0.05$.

Table 2 Normalized cyto-nuclear linkage disequilibrium (D') between 14 nuclear loci and a SNP in the mitochondrial gene *atp1* found in populations of *Silene latifolia* sampled in 1993 and 2008, as well as a test for heterogeneity among pairs of correlation coefficients, or D' , between years

Locus	D'_{1993}	D'_{2008}	D'_{pooled}	t-value (1993 vs 2008)
SL_eSSR04	0.1872	0.1476	0.1737*	0.347
SL_eSSR06	-0.1061	-0.2342*	-0.1751**	1.127
SL_eSSR09	0.1813	0.2412**	0.2161**	0.535
SL_eSSR12	0.4643***	0.2382**	0.3243***	2.216
SL_eSSR16	0.4328***	0.2303*	0.3258***	1.951
SL_eSSR20	-0.1964	-0.6316	-0.3721	8.042***
SL_eSSR29	0.1857	0.3529**	0.2841**	1.542
SL_eSSR30	0.0892	0.0897	0.0691	0.004
slat_18	0.0379	0.0879	0.071	0.428
slat_32	-0.2459**	-0.3709***	-0.3161***	1.180
slat_33	0.6211***	0.3888***	0.4868***	2.698*
slat_48	0.172	0.2939***	0.2415***	1.101
slat_72	0.375***	-0.366**	0.0193	6.634***
slat_85	-0.424***	0.0045	-0.1492*	3.897***

Statistical significance is indicated by the symbols * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. The t-statistic has infinite df, and so can be considered significant at 1.96.

Figure 1 Standardized cyto-nuclear LD (D') between 14 nuclear loci and a SNP in the mitochondrial gene *atp1* found in populations of *S. latifolia* sampled in 1993 and 2008 with a well-characterized metapopulation located in Giles County, VA, USA. Values of D' can range from -1.0 to 1.0. Within the sampled plant populations, D' was found to range from -0.424 to 0.6211 in 1993 and -0.6316 to 0.3888 in 2008. There was a general trend for an overall decrease in D' between sampling periods, though only slat85 showed a significant (positive) change in overall cyto-nuclear D' based upon our t-test of samples (* $P < 0.05$ after Benjamini-Hochberg correction). slat33, slat72 and SL_eSSR20 all showed a significant decrease in D' .

values vary by more than an order of magnitude, ranging from 0.019 for locus *slat72* to 0.487 for locus *slat33*. The relative rankings of locus-specific D' values were consistent between years (see Figure 1),

yielding a between-year Spearman Rank Correlation value (Sokal and Rohlf, 2012) of $r_s = 0.653$ ($P = 0.01$). This view of sample-to-sample temporal consistency is reinforced by the fact that only 5 of 14 locus-specific year-to-year heterogeneity tests were significant (Table 2). Given the year-to-year consistency in the locus-specific D' values, it seems warranted to test whether the wide range of such values noted above does, in fact, represent statistically significant among-locus heterogeneity. Application of the method of Sokal and Rohlf (2012) to the 14 pooled absolute D' values indicates statistically significant heterogeneity in locus-specific values ($\chi^2 = 78.59$, $df = 13$, $P < 0.001$).

DISCUSSION

The results demonstrate that cyto-nuclear associations can be highly variable depending on which nuclear locus is considered, and can average ~25% of the theoretical maximum as defined by allele frequencies. Furthermore, this locus-specific pattern persists across independent samples separated by more than seven generations. This raises three questions. (1) What is the reason certain loci are associated with different cytoplasmic backgrounds while others are not? (2) Why is this heterogeneity in locus-specific cyto-nuclear LD temporally stable? (3) What are the evolutionary consequences of these cyto-nuclear disequilibria?

Given that the nuclear and cytoplasmic genomes are not physically linked, in purely outcrossing species like *S. latifolia*, the accumulation of LD must be due either to epistatic selection or neutral demographic processes associated with the history of population structure (for example, founder effects). These founder effects could include ongoing local events that occur regularly in metapopulations (McCauley, 1994; McCauley *et al.*, 1996; McCauley, 1997; Keller *et al.*, 2012), or historical processes including the species post-glacial history of range expansion (Taylor and Keller, 2007).

S. latifolia has a history of post-glacial expansion in the species native range that would favor the persistence of allelic associations. *S. latifolia* was introduced to North America relatively recently, ca. 200 years ago (Taylor and Keller, 2007). Following multiple, likely separate, introductions to both the Eastern and Western coasts of North America, the introduced range of *S. latifolia* expanded rapidly (Taylor and Keller, 2007; Keller *et al.*, 2009). Studies have concluded that during the invasion process, *S. latifolia* has maintained genetic diversity, though the distribution of genetic variation at various hierarchical levels has become reorganized (Keller *et al.*, 2012). Particularly striking is the decrease in among-regional scale allelic differentiation (F_{RT}) in North America, as compared with the species' native range. However, at the scale of local populations, the level of genetic structure (F_{PT}) is much more comparable between ranges ($F_{PT,Nuclear-EU} = 0.147$, $F_{PT,Nuclear-NA} = 0.131$; $F_{PT,cpDNA-EU} = 0.498$, $F_{PT,cpDNA-NA} = 0.382$). This scenario presents the hypothesis that ancestral population structure in the native range, and the LD that would result, is incompletely dissolved by the incomplete admixture occurring in the introduced range. Accordingly, the observed cyto-nuclear associations seen in the species' native range remain significant, with some evidence of a significant shift in cyto-nuclear associations, suggesting the action of admixture process to mix nuclear and organelle constituents (Keller *et al.*, 2012). This admixture, however, is not a uniform process, and the fact remains that many ancestral allelic associations may persist. Given the results of Keller *et al.* (2012), the shifting cyto-nuclear associations detected here could be a local reflection of the admixture among European lineages following relatively recent establishment in North America. Though the focal organellar genome differed between the prior

studies and our study (cpDNA versus mtDNA), the high likelihood that each is inherited maternally should make their histories congruent, or nearly so.

The present study focuses on nuclear- and cyto-nuclear LD at a finer spatial scale than in previous studies of *S. latifolia*. As the microsatellite loci and organelle locus utilized in the present study are assumed to be unlinked and neutral with respect to fitness, the observed measures of F_{IT} (Table 1), nuclear–nuclear LD (Table 1; Supplementary Tables 1 and 2) and cyto-nuclear LD likely reflect random sampling processes, such as that arise within spatio-temporally distributed metapopulations (Hanski and Gaggiotti, 2004). Further analyses of the sort described will likely corroborate the observed patterns, given observed levels of population subdivision (Edelaar *et al.*, 2011), though further research will be required to pinpoint the effect of individual spatio-temporal characters in determining the magnitude of cyto-nuclear LD (see below).

The finding that the overall level of cyto-nuclear LD has persisted over more than seven generations, even as some of the underlying allelic associations have shifted, could result from local population structure. Under panmixis and selective neutrality, one would expect cyto-nuclear LD to decay by $\sim 50\%$ each generation. Any deviation from panmixis will slow this process, as non-random mating would limit the opportunity for cyto-nuclear genotypic mixing. The collection of individuals contributing to this data set showed a marked deviation from panmixia as evidenced by a F_{IT} value of 0.35. This is not surprising, given that prior studies of *S. latifolia* (= *S. alba*) in this region of Virginia detected significant population structure as well ($F_{ST}=0.20$) when local patches were used to define populations (McCauley, 1994; McCauley *et al.*, 1995). Further studies revealed additional structuring within those patches (McCauley *et al.*, 1996; McCauley, 1997). As the present data set does not account for the fine-scale arrangement of individuals within patches (the sample size per patch is too small for meaningful estimates of very local population structure), our high value of F_{IT} probably reflects within and among-patch divergence. Given that *S. latifolia* is dioecious and hence an obligate outcrosser, spatial structuring rather than self-fertilization must be responsible for these patterns.

In contrast to the moderate level of nuclear F_{ST} found in previous studies, values of F_{ST} based on cpDNA markers were very high—approximately 0.65 when defined at the patch level (McCauley, 1995). Spatial sub-structuring of cpDNA was also found within populations (McCauley *et al.*, 1996). Thus, the movement of maternally inherited cytoplasmic genes (in seeds only) must be considerably more restricted than the movement of nuclear genes (in seeds and pollen) (McCauley, 1997).

An additional feature of *S. latifolia* populations is that local demes in this region of Virginia undergo frequent turnover and that these metapopulation dynamics influence population genetic structure (McCauley *et al.*, 1995). This could affect the persistence of cyto-nuclear LD depending on the degree of mixing that accompanies local colonization events. Slatkin (1977) modeled two modes of colonization—the ‘propagule pool’ mode in which all k individuals contributing to a given colonization event are drawn from the same source, and the ‘migrant pool’ mode in which a group of k colonists are drawn from a genetically representative sample of sources. Whitlock and McCauley (1990) define ϕ as the probability that two alleles in a newly formed population were drawn from the same source population ($\phi=1$ for propagule pool colonization). While these models were developed to consider the effect of colonization on F_{ST} (for a given number of colonists, k , F_{ST} tends to increase as ϕ increases), the same logic will apply to allelic associations.

Colonization events in which ϕ approaches zero (migrant pool) would provide genetic mixing that would enhance the decay of allelic associations, whereas ancestral disequilibria are expected to persist under a propagule pool model. McCauley *et al.* (1995) showed that local population structure was a consequence of these founding events. Moreover, they estimated that ϕ was on the order of 0.80, meaning that colonization events in this *S. latifolia* metapopulation provide relatively little opportunity for the breakup of allelic associations. Taken together, the local spatial structuring of the nuclear and cytoplasmic genomes could clearly slow the rate of decay of nuclear and cyto-nuclear LD relative to an expectation based on panmixis. Finally, it is important to point out that Slatkin’s migrant pools and propagule pools are themselves oversimplifications of any natural situation. For example, even if the sources of founders are diverse so that old associations may dissolve after admixture, if the number of founders are few then new associations may form. This dynamic would generate the type of result that we observed, where the specific associations are shifting, but statistical associations of some form persist.

As an example of this latter process, we present a simple model. Within this model, we assume two bi-allelic loci. Let A and B be the alleles at the first locus, 1 and 2 be the alleles at the second locus. Now consider four differentiated source populations, fixed for genotypes A1, A2, B1 and B2, respectively. Each contributes 20, 60, 60 and 10 colonists, respectively, to a newly colonized study site. Under this scenario, A and 1 would be the most common alleles in this newly mixed population, yet a negative LD for the A1 genotype would be detected. Importantly, under this entirely neutral drift-like process, significant negative or positive LD values will be generated.

A persistent association among nuclear and organellar loci could be important for understanding many evolutionary processes. Persistent cyto-nuclear associations would result from strong cyto-nuclear epistasis for fitness. A well-known example of this is the interaction between mitochondrial cytoplasmic male sterility and nuclear restorer sex-determining loci seen in many cases of gynodioecy (McCauley and Bailey, 2009). The accumulation of positive epistatic interactions among nuclear and cytoplasmic genomes could be responsible for cyto-nuclear incompatibilities when those interactions are disrupted in crosses among lineages (Moyle *et al.*, 2004). This could have implications in the process of speciation (Sambatti *et al.*, 2008). Persistent cyto-nuclear associations have implications for the co-evolution of the two genomes, which could favor the transfer of genes from the mitochondrial to nuclear genome (Brandvain and Wade, 2009).

Theoretical models developed to predict when selection can effectively act on allelic associations among interacting genomes have focused on the term θ , the joint probability that a pair of genes on each respective genome is identical-by-descent (Wade and Goodnight, 2006). As the magnitude of θ increases, the degree to which cyto-nuclear gene combinations are inherited together in transmission from parents to offspring also increases (Wade and Goodnight, 2006). Many models have explored how hybridization and patterns of non-random mating lead to different patterns of θ and the resulting cyto-nuclear disequilibria (Basten and Asmussen, 1997). Additionally, population structure and the resulting increase in bi-parental inbreeding will enhance θ , resulting in increased cyto-nuclear disequilibria, and hence the potential for selection on cyto-nuclear interactions (Brandvain and Wade, 2009).

The results presented here suggest that the magnitude of cyto-nuclear LD necessary for selection to act on cyto-nuclear interactions may be found in these *S. latifolia* populations, at least

for some nuclear/organelle combinations. These associations likely accumulated from the combined influence of invasion history and current metapopulation structure. Importantly, some of our data suggest that allelic associations could be transient, reduced by admixture and regenerated by founder effect. If that were a general result, then an instantaneous estimate of cyto-nuclear LD may over estimate the potential for longer-term co-evolutionary interactions among organellar genomes.

Few empirical studies have quantified the magnitude of intra-species cyto-nuclear LD (Latta *et al.*, 2001), and none have quantified how the magnitude and among-marker variance in cyto-nuclear associations have changed over time. Estimating the magnitude and variance in allelic associations over time, and identifying the processes (both neutral and selective) that generate these associations would contribute to our understanding of the causes and consequences of co-evolution among eukaryotic organelles.

DATA ARCHIVING

Data deposited in the Dryad repository: doi:10.5061/dryad.7c730.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Financial support for this work was provided by NSF grants DEB no. 1050331 and 9221175 to DEM. and NSF DEB no. 0919335 to DRT.

Abdoulaye D, Acevedo I, Adebayo AA, Behrmann-Godel J, Benjamin RC, Bock DG *et al.* (2010). Permanent genetic resources added to Molecular Ecology resources database 1 August 2009–30 September 2009. *Mol Ecol Res* **10**: 232–236.

Andras JP, Ebert D (2013). A novel approach to parasite population genetics: experimental infection reveals geographic differentiation, recombination and host-mediated population structure in *Pasteuria ramosa*, a bacterial parasite of *Daphnia*. *Mol Ecol* **22**: 972–986.

Arnold ML (2006). *Evolution Through Genetic Exchange*. Oxford University Press: Oxford, UK.

Asmussen MA, Arnold J, Avise JC (1989). The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics* **122**: 923–934.

Asmussen MA, Schnabel A (1991). Comparative effects of pollen and seed migration on the cytonuclear structure of plant-populations. 1. Maternal cytoplasmic inheritance. *Genetics* **128**: 639–654.

Baker WK (1975). Linkage disequilibrium over space and time in natural populations of *Drosophila montana*. *Proc Natl Acad Sci USA* **72**: 4095–4099.

Barrière A, Félix MA (2007). Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. *Genetics* **176**: 999–1011.

Basten CJ, Asmussen MA (1997). The exact test for cytonuclear disequilibrium. *Genetics* **146**: 1165–1171.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Stat Method)* **57**: 289–300.

Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D *et al.* (2009). *Silene* as a model system in ecology and evolution. *Heredity* **103**: 5–14.

Brandvain Y, Wade MJ (2009). The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. *Genetics* **182**: 1129–1139.

Datta S, Fu YX, Arnold J (1996). Dynamics and equilibrium behavior of cytonuclear disequilibria under genetic drift, mutation, and migration. *Theor Popul Biol* **50**: 298–324.

Edelaar P, Burraco P, Gomez-Mestre I (2011). Comparisons between Qst and Fst—how wrong have we been? *Mol Ecol* **20**: 4830–4839.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**: 564–567.

Gaunt TR, Rodriguez S, Zapata C, Day INM (2006). MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* **7**: 227.

Giles B, Goudet J (1997). Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *Am Nat* **149**: 507–526.

Goudet J (2002). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from <http://www2.unil.ch/popgen/softwares/fstat.htm>.

Hanski I, Gaggiotti O (2004). Metapopulation biology: past, present and future. In: Hanski I, Gaggiotti O (eds) *Metapopulation Ecology, Genetics and Evolution*. Academic Press: San Diego, CA, USA, pp 3–22.

Hedrick PW (1987). Gametic disequilibrium measures—proceed with caution. *Genetics* **117**: 331–341.

Keller SR, Gilbert KJ, Fields PD, Taylor DR (2012). Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Mol Ecol* **21**: 4721–4734.

Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR (2009). Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytol* **183**: 678–690.

Langley CH, Ito K, Voelker RA (1977). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. *Genetics* **86**: 447–454.

Latta RG, Linhart YB, Mitton JB (2001). Cytonuclear disequilibrium and genetic drift in a natural population of ponderosa pine. *Genetics* **158**: 843–850.

Lewontin RC (1964). The Interaction of selection and linkage. I. General Considerations; heterotic models. *Genetics* **49**: 49–67.

Li MH, Merila J (2010). Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**: 600–610.

Matschiner M, Salzburger W (2009). TANDEM: integrating automated allele binning into genomics and genomics workflows. *Bioinformatics* **25**: 1982–1983.

McCauley DE (1994). Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: implications for studies of gene flow in plants. *Proc Natl Acad Sci USA* **91**: 8127–8131.

McCauley DE (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends Ecol Evol* **10**: 198–202.

McCauley DE (1997). The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *J Hered* **88**: 257–263.

McCauley DE, Bailey MF (2009). Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Ann Bot* **104**: 611–620.

McCauley DE, Ellis JR (2008). Recombination and linkage disequilibrium among mitochondrial genes in structured populations of the gynodioecious plant *Silene vulgaris*. *Evolution* **62**: 823–832.

McCauley DE, Raveill J, Antonovics J (1995). Local founding events as determinants of genetic structure in a plant metapopulation. *Heredity* **75**: 630–636.

McCauley DE, Stevens J, Peroni P, Raveill J (1996). The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *Am J Bot* **83**: 727–731.

Meirmans PG, Van Tienderen PH (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* **4**: 792–794.

Moccia MD, Oger-Desfeux C, Marais GA, Widmer A (2009). A white campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics* **10**: 243.

Moyle LC, Olson MS, Tiffin P (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution* **58**: 1195–1208.

Nei M, Li WH (1973). Linkage disequilibrium in subdivided populations. *Genetics* **75**: 213–219.

Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2008). Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol Lett* **11**: 1082–1091.

Slate J, Pemberton JM (2007). Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *J Evol Biol* **20**: 1415–1427.

Slatkin M (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol* **12**: 253–262.

Slatkin M (1994). Linkage disequilibrium in growing and stable populations. *Genetics* **137**: 331–336.

Slatkin M (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**: 477–485.

Sokal RR, Rohlf FJ (2012). *Biometry: The Principles and Practice of Statistics in Biological Research*, 4th edn WH Freeman and Co.: New York, NY, USA.

Stinchcombe JR, Hoekstra HE (2007). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.

Taylor DR, Keller SR (2007). Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* **61**: 334–345.

Wade M, Goodnight C (2006). Cyto-nuclear epistasis: two-locus random genetic drift in hermaphroditic and dioecious species. *Evolution* **60**: 643–659.

Whitlock MC, McCauley DE (1990). Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717–1724.

Zapata C (2011). On the uses and applications of the most commonly used measures of linkage disequilibrium from the comparative analysis of their statistical properties. *Human Heredity* **71**: 186–195.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)