CYTO-NUCLEAR LINKAGE DISEQUILIBRIUM RESULTING FROM ADMIXTURE

Peter D. Fields[1,2,4], David E. McCauley[3,] and Douglas R. Taylor[1]

[1]Department of Biology, University of Virginia, PO Box 400328, Charlottesville, VA 22904-4328

[2] Current Address: University of Basel, Zoological Institute, Vesalgasse 1, Basel, CH-4051, Switzerland

[3] Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235

[4] Corresponding author: Email: peter.fields@unibas.ch

Running Head:

Keywords:

ABSTRACT

Previous studies of North American populations of the invasive plant *Silene latifolia* showed significant cyto-nuclear linkage disequilibrium (CNLD) between SNP variants of a mitochondrial gene (*atp*1) and the most common allele at nuclear microsatellite loci. Fields *et al*. (2014) hypothesized that this CNLD arose partially as a consequence of admixture that occurred during the colonization of North American (NA) populations of *S latifolia* via seed dispersal from genetically differentiated European populations that represent a portion of the native range of this species. In order to evaluate the plausibility of the admixture hypothesis, as opposed to metapopulation processes alone, we estimated CNLD for these same loci using data collected from eastern (EEU) and western (WEU) European populations of *S. latifolia* known to be genetically differentiated and likely sources of the spread of the study species to North America. We show that the CNLD found previously in NA populations of *S. latifolia* can be attributed to admixture of the previously isolated European demes coupled with decay since that time. Our applied framework allows the separation of the forces generating and dissolving statistical associations between alleles in cytoplasmic organelles and the nuclear genome and may thus be of utility in the study of plant or animal microbiomes.

INTRODUCTION

The magnitude and distribution of genetic variation across the recently established range of an invasive species is a reflection of the demographic history of demes from the native range, colonization events, and other evolutionary processes such as natural selection and genetic drift within the introduced range (Taylor & Keller, 2007, Dlugosch & Parker, 2008, Hufbauer, 2008, Calsbeek et al., 2011, Colautti & Barrett, 2011, Keller et al., 2014). For our purposes genetic variation can be defined as the degree of polymorphism at some number of loci as measured in aggregate across the new range, or as measured both within local populations as well as among those populations (i.e., considering the geographic population genetic structure). Measurements of variation could be extended beyond the standard locus-by-locus variety to include the distribution within and among populations of various multi-locus genotypes. These could include evaluation of non-random associations of alleles across loci (linkage disequilibrium or LD), again considering the genetic composition of all populations in aggregate or on a population-by-population basis.

In order to understand the determinants of genetic variation described above, it would be useful to consider mechanisms of short and long-range gene dispersal, along with the influence of the mating system, on multi-locus genotypic diversity. Note that in many species of seed plants most genes have the potential to disperse in both seeds and pollen. However, because the chloroplast and mitochondrial genomes tend to be extensively (or entirely) maternally inherited (McCauley, 2013) the movement of pollen would appear to be a more important

63 mechanism for long distance dispersal of nuclear genes as compared to that of
64 organellar genes. This distinction would appear to be especially relevant to the
65 emigration of genetic diversity during the early stages of an invasion since a new
66 population cannot be founded initially by pollen alone and thus requires seed
67 dispersal (Baker & Stebbins, 1965). On the other hand, long distance seed dispersal
68 can transport cyto-nuclear genotype combinations that exist in the source
69 populations (Hu, 2008). That is, when cyto-nuclear linkage disequilibrium (CNLD) is
70 found in an invasive plant species one must consider whether that CNLD is a
71 reflection of the genetic composition of the populations from which colonists were
72 drawn. The possibilities include that the CNLD occurred in some populations in the
73 native range, or that it was created as colonists from genetically differentiated
74 populations in the native range underwent admixture at or after the time of the
75 invasion. However, it is known from theory that CNLD found in neutral variation
76 should decline by 50% per generation in panmictic populations so CNLD persistence
77 is problematic whatever the origin (Fields et al., 2014, Hu, 2008). Note also, though,
78 that numerous species of plants are predominant selfers, and that recurrent self-
79 fertilization limits the opportunity for the mixing of alleles between genomes. In
80 addition to helping to reveal the history of populations regarding patterns of gene
81 flow and mating system, study of CNLD can also be useful when considering the role
82 of cyto-nuclear co-evolution, including selective phenomenon leading to
83 coadaptation or conflict between the genomes (Fields et al., 2014, Sloan, 2015, Bock
84 et al., 2014) and the transfer of genetic information between genomes (Sloan, 2015).

85 In a recent study of the plant *Silene latifolia*, Fields et al. (2014) identified
86 high levels of cyto-nuclear linkage disequilibrium (CNLD) between variants of a SNP
87 that occurs within the mitochondrial gene *atp1* and alleles at a number of nuclear
88 microsatellite loci. Samples were taken from populations located in southwestern
89 Virginia, USA, within what is assumed to be the eastern portion of the invasive range
90 of *S. latifolia* in North America (Keller et al., 2012). A comparison to genotypes found
91 in the same *S. latifolia* populations sampled 20 years (~ 7 generations) prior to the
92 more recent samples revealed only modest change in mean CNLD through time.
93 Fields *et al.* (2014) speculated that the CNLD that was observed could be traced to
94 1) metapopulation dynamics, in particular founder effects associated with a
95 propagule pool (Slatkin, 1977, Wade & McCauley, 1988, Whitlock & McCauley, 1990)
96 colonization dynamic, and 2) the colonization process by which the European native
97 *S. latifolia* became established in this region of the eastern U.S. via seed dispersal.
98 The relative temporal stability of the observed CNLD could be due to the fact that *S.*
99 *latifolia* displays a high degree of local population structure in this area (McCauley,
100 1994) and that this would limit opportunities for the outbreeding needed for the
101 erosion of any CNLD that arose when those populations were established. A link
102 between CNLD and the establishment of populations would be especially likely if the
103 various European populations contributing colonists (either directly or indirectly
104 via dispersal from other recently established North American populations)
105 displayed some degree of cyto-nuclear differentiation that could contribute to CNLD,
106 especially with admixture.

107 Extensive recent sampling across much of the putative native range of *S.*
108 *latifolia* in Europe has revealed geographic genetic structuring that is especially
109 noticeable when populations from eastern and western Europe are compared
110 (Keller et al., 2012, Keller et al., 2014, Taylor & Keller, 2007). Here we investigate
111 the plausibility that the CNLD in North American populations found by Fields et al.
112 (2014) could be a consequence of the recent establishment of *S. latifolia* in North
113 America and admixture between divergent European populations during
114 hypothetical invasion events. We find that different invasion scenarios, combined
115 with different forms of cyto-nuclear population differentiation in the native range,
116 could have enhanced or diminished the CNLD observed in the invasive range.
117 Application of cyto-nuclear genotyping, in addition to nuclear genotyping alone,
118 provides additional insights for the origin of statistical associations of genetic
119 variation among genomic regions, and the relative roles of historical demography
120 and ongoing metapopulation processes.

121 ***Theoretical expectations for shifts in CNLD***

122 Consider a population of individuals that are characterized at two loci, one
123 cytoplasmic (C) and one nuclear (N) with two alleles at each locus (C_1 & C_2; N_1 & N_2).
124 This population could be described according to the frequency (p) of one of the
125 alleles at each locus (i.e. $p_{C,1}$ and $p_{N,1}$) and by the frequency of the cytonuclear
126 genotypes (i.e. $p_{C,N,1\ etc}$). The disequilibrium between C_1 & N_1, $D_{C\text{-}N}$, could be defined
127 in this population as $D_{C\text{-}N} = [p_{C,N,1} - ((p_{C,1})(p_{N,1}))]$. Thus, $D_{C\text{-}N}$ determines whether two
128 alleles are associated within individuals, more (positive) or less (negative) often
129 than random and is mathematically equivalent to the covariance between alleles.

130 Note that the equation for LD is closely related to a covariance (symbolized as D)
131 with maximum values constrained by the gene frequencies in the sample. Another
132 standardized measure of CNLD (D') is also commonly reported (e.g. in Fields et al.
133 (2014)). This standardization constrains values of D' to fall between -1 and 1
134 (negative and perfect association of alleles, respectively).

135 Now consider $p_{C,1}$, $p_{N,1}$ and $p_{C,N,1}$ for each of k = 2 or more populations
136 individually. From this information, the overall CNLD value could be calculated in
137 two ways. First, $D_{C\text{-}N}$ can be calculated for each of the k populations, and assuming
138 the populations are identical in size, these values could be averaged to obtain $\bar{D}_{C\text{-}N}$.
139 Alternatively, individual genotypes could be pooled across the k populations and
140 used to calculate values of $p_{C,1}$, $p_{N,1,}$, and $p_{C,N,1}$ $p_{C,1}p_{N,1}$ and $D^{T}_{C\text{-}N}$ based on the pooled
141 total set of individual genotypes.

142 Cytonuclear disequilibrium, and LD in general, may arise from several
143 sources or by processes operating at different spatial scales. First, at a larger spatial
144 scale, CNLD will arise naturally as a result of population genetic divergence at both
145 nuclear and cytoplasmic loci. Individual populations may also display variation in
146 the population-specific values of $D_{\text{C-N}}$ according to how the alleles at the nuclear and
147 cytoplasmic loci are organized into two-locus genotypes in each population, which
148 could be a function of a variety of forces operating within populations (e.g. mating
149 structure, demographic history, selection, etc.).

150 One might predict how the cyto-nuclear genetic structure of a set of *k*

151 populations determines $\mathbf{D^T_{C\text{-}N}}$ in a population originating from mixing individuals

152 from those populations. It can be shown that

153 $$D^T_{C\text{-}N} = [\bar{D}_{C\text{-}N} + Cov(p_{C,1}, p_{N,1})], \quad (1)$$

154 where $Cov(p_{C,1}, p_{N,1})$ is the covariance of the *k* population-specific cytoplasmic and

155 nuclear allele frequencies across populations contributing equally to the mixture (p.

156 550, Hedrick, 2011). Note, this equation does not apply directly to *D'*.

157 Figure 2 illustrates the application of this equation to several hypothetical

158 sets of *k* = 2 source populations. From this equation it can be seen that mixing

159 individuals from more than one source can result in a $D^T_{\text{C-N}}$ in the new population

160 higher than the average $D^T_{\text{C-N}}$ in the source populations if $Cov(p_{C,1}, p_{N,1}) > 0$ is

161 (Figure 2, scenario 1). $Cov(p_{C,1}, p_{N,1})$ would be positive if certain combinations of

162 cytoplasmic and nuclear alleles coexist in the same populations more often than a

163 random association across populations. Mixing can diminish $D_{\text{C-N}}$ if $Cov(p_{C,1}, p_{N,1})$ is

164 negative (Figure 2, scenario 2; scenario 3), as would be the case if certain cyto-

165 nuclear alleles co-occur less often than random in the source populations. It can

166 have no effect on $D^T_{\text{C-N}}$ as compared to average population-specific CNLD (ignoring

167 stochastic sampling) if $Cov(p_{C,1}, p_{N,1}) = 0$, as (Figure 2 scenario 4; scenario 5).

168 $Cov(p_{C,1}, p_{N,1}) = 0$ would be the case if the nature of cyto-nuclear associations is

169 consistent across populations.

170 ***CNLD in the European range of S. latifolia***

171 Samples of *S. latifolia* were collected from 85 individuals distributed across
172 12 populations in the species' native European range (Figure 1), a subset of the total
173 sample used by Keller et al. (2012). Genomic DNA was extracted as described in
174 Keller et al. (2012) using methods adapted from Edwards et al. (1991) and Slotta et
175 al. (2008). Individuals were genotyped at 14 microsatellite markers derived from *S.*
176 *latifolia* (slat18, slat32, slat33, slat48, slat72, slat85) described by Molecular Ecology
177 Resources Primer Development et al. (2010), and (SL_eSSR04, SL_eSSR06,
178 SL_eSSR09, SL_eSSR12, SL_eSS16, SL_eSSR20, SL_eSSR29, SL_eSSR30) described by
179 Moccia et al. (2009). Fragment analysis was conducted using an ABI 313xl
180 automated sequencer, with genotypes determined using GENEMAPPER v.3.0
181 software (Applied Biosystems) and binned using the program TANDEM (Matschiner
182 & Salzburger, 2009).
183 We followed the methods described by Keller et al. (2012) in order to
184 distinguish the two distinct, post-glacial demes in *S. latifolia*'s native range. Briefly,
185 individuals are assigned to Eastern and Western Europe (EEU and WEU,
186 respectively) based upon a suture zone running north from the Alps to the North
187 Sea, and has been described previously as different routes of expansion for *S.*
188 *latifolia* out of distinct Mediterranean refugia following Holocene warming (Hewitt,
189 1996, Petit et al., 2003, Taylor & Keller, 2007, Keller et al., 2009, Keller et al., 2012).
190 Using these *a priori* expectations, Keller et al. (2012) applied Bayesian clustering
191 using the program STRUCTURE v.2.2 (Pritchard et al., 2000, Falush et al., 2003),
192 assuming an admixture model with allele frequencies correlated and remaining
193 parameters set as a default over 10 independent runs with 200,000 MCMC

194 iterations and a burn-in period of 50,000 iterations. Here we utilize cluster
195 assignments assuming K=2 representing EEU and WEU, respectively (Keller et al.,
196 2012, Taylor & Keller, 2007).

197 In order to estimate CNLD, we followed the approach of Fields et al. (2014).
198 The same genomic samples used for the microsatellite analysis were assayed for a
199 SNP in the mitochondrial gene *atp1* using a PCR/RFLP method (McCauley & Ellis,
200 2008, Fields et al., 2014). PCR products were digested with *AluI* (New England
201 Biolabs), electrophoresed on 4% Metaphor agarose gel (Lonza Inc, Rockland, ME,
202 USA), and stained with ethidium bromide in order to resolve the two possible
203 alleles.

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

210 We estimated nuclear-nuclear LD (hereafter, nuclear LD) following the
211 methods described in Fields et al. (2014). Statistical significance of nuclear LD
212 between pairs of loci for post-glacial demes, as well as within the full European
213 sample, was determined using a Monte-Carlo approximation of Fisher's exact test
214 implemented in the software Arlequin (Excoffier & Lischer, 2010), which assumes a
215 null hypothesis of random allelic assortment. Arlequin uses a Markov chain
216 extension of Fisher's exact test for *RxC* contingency tables (Slatkin, 1994, Li &

217 Merila, 2010). 100,000 alternative tables were explored by the Markov chain (Slate
218 & Pemberton, 2007, Li & Merila, 2010).

219 We follow the methods described in Fields et al. (2014) to estimate cyto-
220 nuclear *D* and *D'*. Briefly, we followed the approach of Basten and Asmussen (1997),
221 using the program CNDm to estimate a standardized estimate an allelic *D* and *D'*
222 between each nuclear locus and atp1 mtDNA locus. CNDm uses a Monte-Carlo
223 approach to approximate Fisher's exact test for RxC contingency tables and tests for
224 significant deviations from the null hypothesis of no allelic association (Basten &
225 Asmussen, 1997). For this analysis, all nuclear loci were treated as bi-allelic by
226 pooling all alleles other than the most common allele in the full European sample
227 into a single composite allele (as in Latta et al., 2001).

228 **RESULTS AND DISCUSSION**

229 Cyto-nuclear genotypes were obtained for a total of 85 European (EU)
230 individuals (52 individuals from the Western European Deme (WEU) and 33
231 individuals from the Eastern European Deme (EEU)). For the *atp1* SNP the most
232 common variant over the full European range occurred at a frequency of 0.535
233 (0.929 in EEU, and 0.254 in WEU). Table 1 presents a summary of global estimates
234 of nuclear genetic diversity, population substructure (F_{IT}) and a summary of nuclear
235 LD for the EEU, WEU, and full EU sample. All microsatellite loci were polymorphic (>
236 1 allele) in both regions, and ranged between 2 and 21 alleles. As in Fields et al.
237 (2014), individual loci showed a large range of H_O (0.061-0.761, mean = 0.38) and
238 H_E (0.134-0.902, mean = 0.63). Average F_{IT} for the whole EU sample was 0.498
239 (0.17-0.94), with all loci exhibiting significant ($P < 0.05$) deviations from panmixia,

240 corroborating deviations between H_O and H_E. The observed F_{IT} was larger than that
241 reported in Fields et al. (2014), which is to be expected given differences in scale
242 and the opportunity for admixture that took place during the invasion of the North
243 American continent (Keller et al., 2012). The average number of loci pairs showing
244 significant nuclear-nuclear LD was approximately equal in WEU and EEU (3.86 and
245 5.71, respectively), though the full sample showed approximately twice as much
246 (9.86). Table 2 presents the observed distribution of *D* and *D'* values for EEU, WEU,
247 and the full EU sample, respectively, for each of 14 nuclear microsatellites and the
248 mitochondrial *atp1* SNP. Similarly to Fields et al. (2014), overall mean levels of LD
249 were consistent, while specific locus values varied widely.

250 The main goal of this study was to ask whether the levels of CNLD seen by
251 Fields et al (2014) in a North American (NA) sample of *S. latifolia* could plausibly
252 have resulted from admixture of colonists from the European range of this species,
253 Given this, it would seem useful to compare values of D and/or D' obtained from
254 North America (NA) to those from Europe. To that end we have included the pooled
255 values of *D'* from Table 2 of Fields *et al.* (2014) in their Table 2, along with
256 associated values of *D* calculated from the same data but not presented in Fields et
257 al. (2014).

258 From our Table 2 it can be seen that the range of *D'* for NA was -.372 to.487
259 (mean = .086). For EU the range was -.536 .to .557 (mean = .043). In a similar
260 comparison of *D* values the range for NA was -.078 to .1042 (mean = .016). For EU
261 the range was -.075 to 0.52 (mean = -.009)

262 Given a fair degree of locus-to-locus variation in the nuclear locus-specific
263 values of *D'* and *D* in EU and NA, an additional method a comparing the similarity of
264 CNLD in EU and NA would be to examine the consistency of these locus-specific
265 values by calculating their product-moment correlation (*r*). That correlation is *r* = -
266 .208 for *D'* and *r* = -0.397 for *D*. Neither *r* value is significantly different from zero
267 under a null hypothesis of *r* = 0 ($p > 0.05$, df = 12 in each case), suggesting drift-like
268 dynamics are responsible for shifting CNLD rather than selective agents.
269 What is the impact of the covariance of cyto-nuclear allele frequencies across
270 EEU and WEU on EU *D* values? Recall from eq. 1 and Fig 1 that represent various
271 "scenarios" that such covariance can either increase or decrease the magnitude of
272 EU D relative to the simple mean of site-specific D values. Inspection of Fig. 3A
273 suggests that all of the scenarios are approximated, depending on the nuclear locus
274 in question. Locus-specific Shifts in CNLD D values range from a shift of .058 (locus
275 SLeSSR12) to a decrease of -.1062 (locus Slat 33) (mean shift in D = -.002).
276 Examination of Table 2 and Figure 3A suggests a large role of differing
277 frequency of the most common cytoplasmic allele, wherein the greater frequency of
278 the allele in EEU leads to low values of *D* for most loci, and a larger range of *D* in
279 WEU. Figure 3B describes a similar pattern for *D'*, wherein the near fixation of one
280 cytoplasmic allele in the EEU deme generates a very strong signal of CNLD across 12
281 of 14 nuclear loci. However, both Figure 3A and 3B imply just how drastically
282 admixture of two highly divergent demes can shift the overall mean and variance of
283 CNLD. Thus, our results are in least partial support of our previous hypothesis that
284 the CNLD found in our North American *S. latifolia* samples could result, in part, from

admixture of colonists from two or more sources in the native range of *S. latifolia* that contributed to the expansion of the range of this species into eastern North America.

CONCLUSION

Linkage disequilibrium, either nuclear-nuclear or cyto-nuclear, provides a useful summary of the demographic processes taking place within and among species. While estimation of nuclear-nuclear LD is common, estimation of CNLD is much less so. This is at least in part due to a much less extensive exploration, both theoretical and empirically, of CNLD. This state seems particularly problematic given the general realization of the role of cyto-nuclear interactions in important evolutionary processes, including local adaptation (e.g. Bock et al., 2014) and speciation (e.g. Trier et al., 2014), as well as co-evolutionary dynamics more generally (Sloan, 2015). Fields et al. (2014) provided a temporal study of the how CNLD will shift over time as a result of metapopulation processes. The present study shows that the incomplete admixture of *S. latifolia* in North America also contributed to CNLD, as well as outlining how allele frequencies in the species native range resolved into present patterns of CNLD in the invaded range. Application of the presented framework, and modifications therein, should allow for a fuller understanding the role of selective and non-selective forces in generating patterns of LD both within and among genomes.

ACKNOWLEGEMENTS

308

309 **REFERENCES**

310 Baker, H. G. & Stebbins, G. L. 1965. *The genetics of colonizing species*. Academic Press,
311 New York.

312 Basten, C. J. & Asmussen, M. A. 1997. The exact test for cytonuclear disequilibria.
313 *Genetics* **146**: 1165-1171.

314 Bock, D. G., Andrew, R. L. & Rieseberg, L. H. 2014. On the adaptive value of
315 cytoplasmic genomes in plants. *Molecular Ecology* **23**: 4899-4911.

316 Calsbeek, B., Lavergne, S., Patel, M. & Molofsky, J. 2011. Comparing the genetic
317 architecture and potential response to selection of invasive and native
318 populations of reed canary grass. *Evolutionary Applications* **4**: 726-735.

319 Colautti, R. I. & Barrett, S. C. H. 2011. POPULATION DIVERGENCE ALONG LINES OF
320 GENETIC VARIANCE AND COVARIANCE IN THE INVASIVE PLANT LYTHRUM
321 SALICARIA IN EASTERN NORTH AMERICA. *Evolution* **65**: 2514-2529.

322 Dlugosch, K. M. & Parker, I. M. 2008. Founding events in species invasions: genetic
323 variation, adaptive evolution, and the role of multiple introductions.
324 *Molecular Ecology* **17**: 431-449.

325 Edwards, K., Johnstone, C. & Thompson, C. 1991. A simple and rapid method for the
326 preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*
327 **19**: 1349-1349.

328 Excoffier, L. & Lischer, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs
329 to perform population genetics analyses under Linux and Windows.
330 *Molecular Ecology Resources* **10**: 564-567.

331 Falush, D., Stephens, M. & Pritchard, J. K. 2003. Inference of population structure
332 using multilocus genotype data: Linked loci and correlated allele frequencies.
333 *Genetics* **164**: 1567-1587.
334 Fields, P. D., McCauley, D. E., McAssey, E. V. & Taylor, D. R. 2014. Patterns of cyto-
335 nuclear linkage disequilibrium in Silene latifolia: genomic heterogeneity and
336 temporal stability. *Heredity* **112**: 99-104.
337 Goudet, J. 2002. FSTAT, a program to estimate and test gene diversities and fixation
338 indices (version 2.9.3.2). *Available from*
339 http://www2.unil.ch/popgen/softwares/fstat.htm.
340 Hedrick, P. W. 2011. *Genetics of Populations,* 4th ed. Jones and Bartlett, Boston.
341 Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in
342 divergence and speciation. *Biological Journal of the Linnean Society* **58**: 247-
343 276.
344 Hu, X.-S. 2008. Barriers to the Spread of Neutral Alleles in the Cytonuclear System.
345 *Evolution* **62**: 2260-2278.
346 Hufbauer, R. A. 2008. Biological invasions: Paradox lost and paradise gained. *Current*
347 *Biology* **18**: R246-R247.
348 Keller, S. R., Fields, P. D., Berardi, A. E. & Taylor, D. R. 2014. Recent admixture
349 generates heterozygosity–fitness correlations during the range expansion of
350 an invading species. *Journal of Evolutionary Biology* **27**: 616-627.
351 Keller, S. R., Gilbert, K. J., Fields, P. D. & Taylor, D. R. 2012. Bayesian inference of a
352 complex invasion history revealed by nuclear and chloroplast genetic

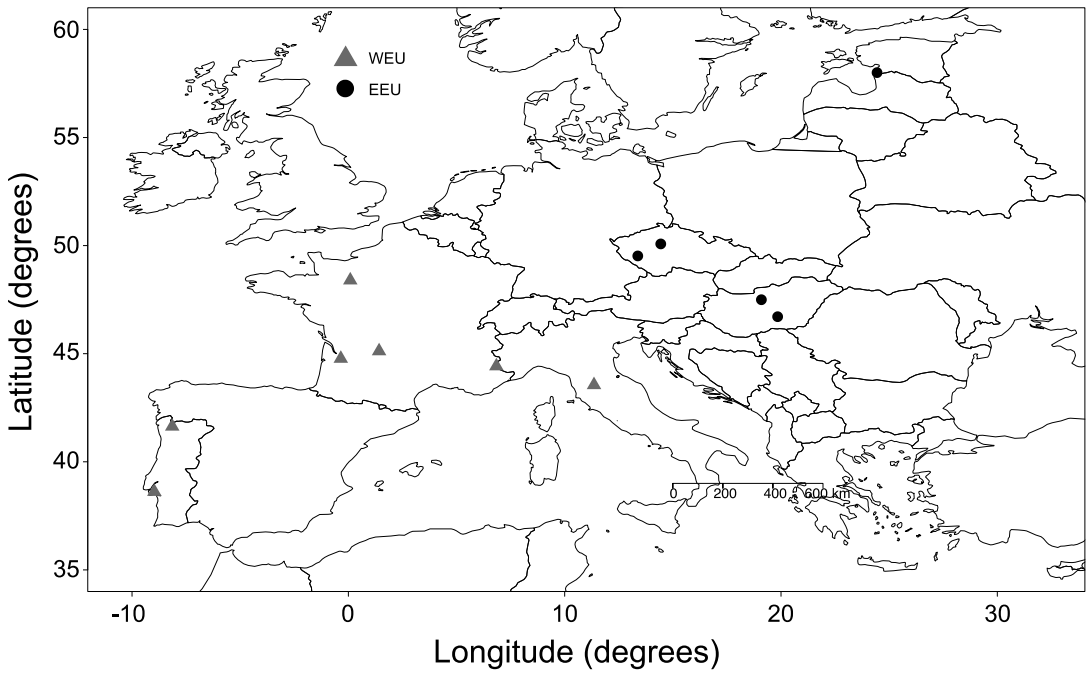
353 diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology* **21**: 4721-
354 4734.
355 Keller, S. R., Sowell, D. R., Neiman, M., Wolfe, L. M. & Taylor, D. R. 2009. Adaptation
356 and colonization history affect the evolution of clines in two introduced
357 species. *New Phytologist* **183**: 678-690.
358 Latta, R. G., Linhart, Y. B. & Mitton, J. B. 2001. Cytonuclear disequilibrium and genetic
359 drift in a natural population of ponderosa pine. *Genetics* **158**: 843-850.
360 Li, M. H. & Merila, J. 2010. Extensive linkage disequilibrium in a wild bird population.
361 *Heredity* **104**: 600-610.
362 Matschiner, M. & Salzburger, W. 2009. TANDEM: integrating automated allele
363 binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982-
364 1983.
365 McCauley, D. E. 1994. Contrasting the distribution of chloroplast DNA and allozyme
366 polymorphism among local populations of Silene alba: implications for
367 studies of gene flow in plants. *Proceedings of the National Academy of*
368 *Sciences of the United States of America* **91**: 8127-8131.
369 McCauley, D. E. 2013. Paternal leakage, heteroplasmy, and the evolution of plant
370 mitochondrial genomes. *New Phytologist* **200**: 966-977.
371 McCauley, D. E. & Ellis, J. R. 2008. Recombination and linkage disequilibrium among
372 mitochondrial genes in structured populations of the gynodioecious plant
373 Silene vulgaris. *Evolution* **62**: 823-832.

374 Meirmans, P. G. & Van Tienderen, P. H. 2004. GENOTYPE and GENODIVE: two
375 programs for the analysis of genetic diversity of asexual organisms.
376 *Molecular Ecology Notes* **4**: 792-794.
377 Moccia, M. D., Oger-Desfeux, C., Marais, G. A. & Widmer, A. 2009. A white campion
378 (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-
379 SSR characterization, transferability, and utility for comparative mapping.
380 *BMC Genomics* **10**: 243.
381 Molecular Ecology Resources Primer Development, C., Abdoullaye, D., Acevedo, I.,
382 Adebayo, A. A., Behrmann-Godel, J., Benjamin, R. C., Bock, D. G., Born, C.,
383 Brouat, C., Caccone, A., Cao, L.-Z., Casado-AmezÚA, P., CatanÉO, J., Correa-
384 Ramirez, M. M., Cristescu, M. E., Dobigny, G., Egbosimba, E. E., Etchberger, L.
385 K., Fan, B. I. N., Fields, P. D., Forcioli, D., Furla, P., Garcia de Leon, F. J., GarcÍA-
386 JimÉNez, R., Gauthier, P., Gergs, R., GonzÁLez, C., Granjon, L., GutiÉRrez-
387 RodrÍGuez, C., Havill, N. P., Helsen, P., Hether, T. D., Hoffman, E. A., Hu, X.,
388 Ingvarsson, P. K., Ishizaki, S., Ji, H., Ji, X. S., Jimenez, M. L., Kapil, R., Karban, R.,
389 Keller, S. R., Kubota, S., Li, S., Li, W., Lim, D. D., Lin, H., Liu, X., Luo, Y.,
390 Machordom, A., Martin, A. P., Matthysen, E., Mazzella, M. N., McGeoch, M. A.,
391 Meng, Z., Nishizawa, M., O'Brien, P., Ohara, M., Ornelas, J. F., Ortu, M. F.,
392 Pedersen, A. B., Preston, L., Ren, Q. I. N., Rothhaupt, K.-O., Sackett, L. C., Sang,
393 Q., Sawyer, G. M., Shiojiri, K., Taylor, D. R., Van Dongen, S., Van Vuuren, B. J.,
394 Vandewoestijne, S., Wang, H., Wang, J. T., Wang, L. E., Xu, X.-L., Yang, G., Yang,
395 Y., Zeng, Y. Q., Zhang, Q.-W., Zhang, Y., Zhao, Y. & Zhou, Y. A. N. 2010.
396 Permanent Genetic Resources added to Molecular Ecology Resources

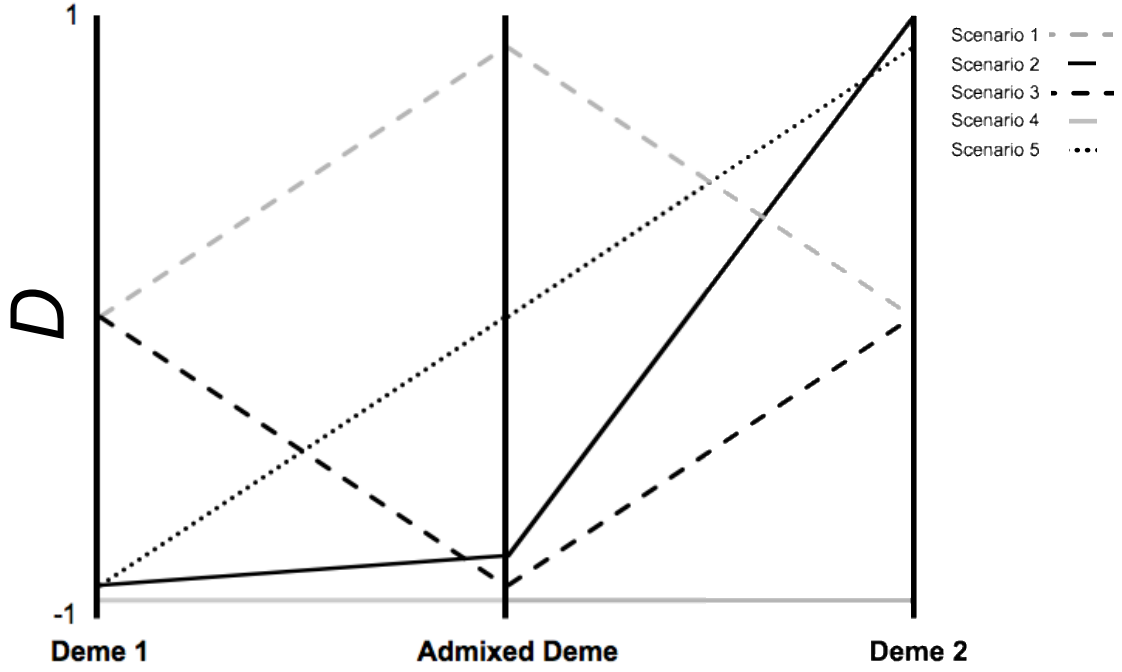
397 Database 1 August 2009–30 September 2009. *Molecular Ecology Resources*
398 **10**: 232-236.
399 Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R.,
400 Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G.,
401 Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S. & Vendramin, G. G.
402 2003. Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity.
403 *Science* **300**: 1563-1565.
404 Pritchard, J., Stephens, M. & Donnelly, P. 2000. Inference of population structure
405 using multilocus genotype data. *Genetics* **155**: 945-959.
406 Slate, J. & Pemberton, J. M. 2007. Admixture and patterns of linkage disequilibrium
407 in a free-living vertebrate population. *Journal of Evolutionary Biology* **20**:
408 1415-1427.
409 Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local
410 extinctions. *Theoretical Population Biology* **12**: 253-262.
411 Slatkin, M. 1994. Linkage disequilibrium in growing and stable populations. *Genetics*
412 **137**: 331-336.
413 Sloan, D. B. 2015. Using plants to elucidate the mechanisms of cytonuclear co-
414 evolution. *New Phytologist* **205**: 1040-1046.
415 Slotta, T. A. B., Brady, L. & Chao, S. 2008. High throughput tissue preparation for
416 large-scale genotyping experiments. *Molecular Ecology Resources* **8**: 83-87.
417 Taylor, D. R. & Keller, S. R. 2007. Historical range expansion determines the
418 phylogenetic diversity introduced during contemporary species invasion.
419 *Evolution* **61**: 334-345.

420 Trier, C. N., Hermansen, J. S., Sætre, G.-P. & Bailey, R. I. 2014. Evidence for Mito-
421 Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian
422 Sparrow and Its Parent Species. *PLoS Genet* **10**: e1004075.
423 Wade, M. J. & McCauley, D. E. 1988. Extinction and recolonization: their effects on
424 the genetic differentiation of local populations. *Evolution* **42**: 995-1005.
425 Whitlock, M. C. & McCauley, D. E. 1990. Some population genetic consequences of
426 colony formation and extinction: genetic correlations within founding
427 groups. *Evolution* **44**: 1717-1724.
428
429

Figure 1. European sampling localities. Grey dots indicate samples from the Western European (WEU) deme and black dots indicate samples from Eastern European (EEU) deme.



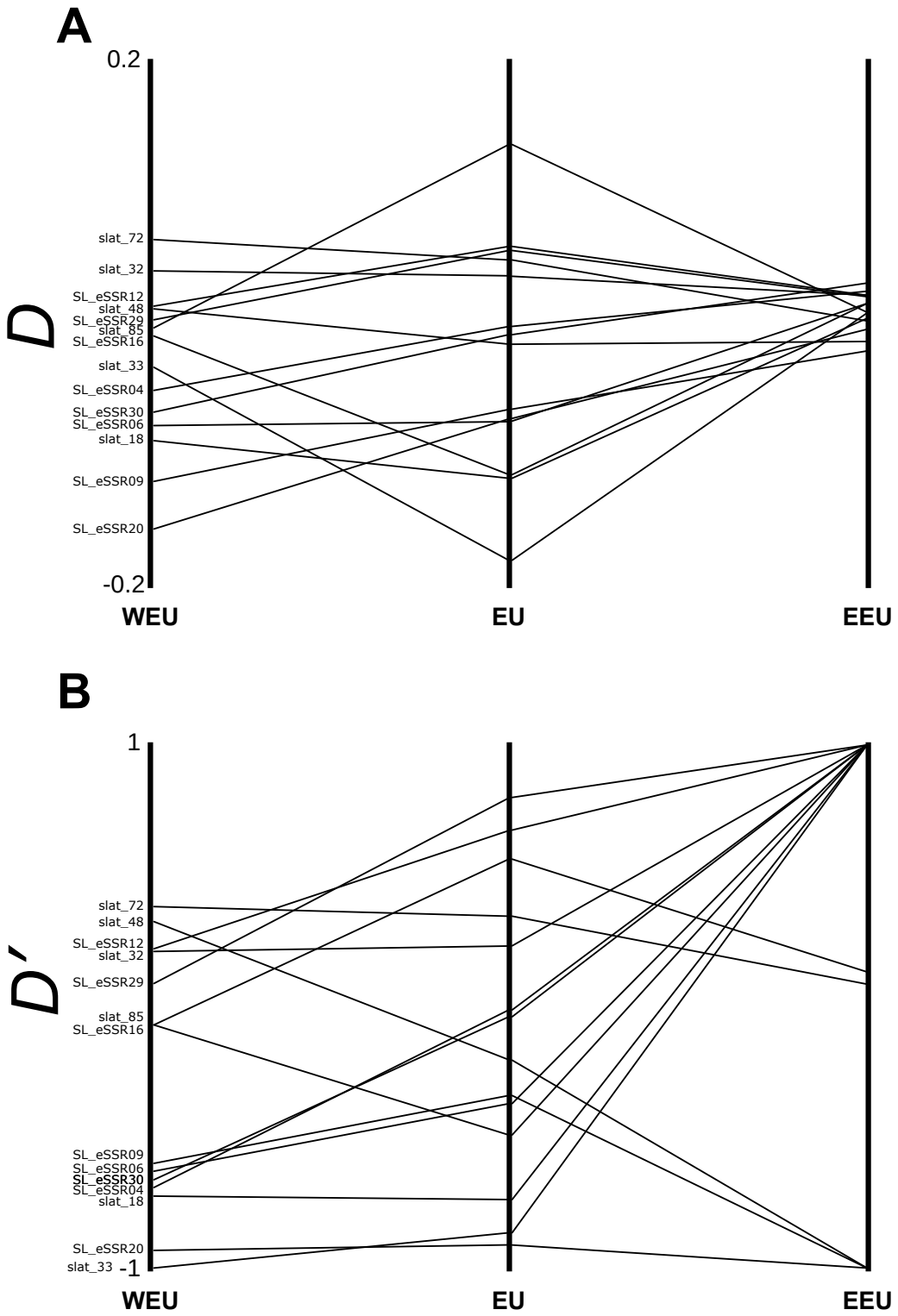
435 **Figure 2.** Hypothetical consequences for CNLD *D* values when populations that
436 differ in their respective values of *D*, due to variable genetic compositions,
437 contribute individuals to form an admixed deme. Five types of mixtures/scenarios
438 are described in the text.



439

440

441 **Figure 3.** Locus by locus CNLD Values of *D* (3A) or *D'* (3B) obtained from single
442 European sources (EEU and WEU) and from pooling samples from those sources
443 (EU). The shift predicted from eq. 1 can be visualized by noting that the D value seen
444 in EU are not always the average of the locus-specific D found in EEU and WEU.
445


A
0.2
D
slat_72
slat_32
SL_eSSR12
slat_48
SL_eSSR29
slat_85
SL_eSSR16
slat_33
SL_eSSR04
SL_eSSR30
SL_eSSR06
slat_18
SL_eSSR09
SL_eSSR20
-0.2
WEU
EU
EEU
B
1
D´
slat_72
slat_48
SL_eSSR12
slat_32
SL_eSSR29
slat_85
SL_eSSR16
SL_eSSR09
SL_eSSR06
SL_eSSR30
SL_eSSR04
slat_18
SL_eSSR20
slat_33
-1
WEU
EU
EEU

446
447

Table 1. Global estimates of genetic diversity, population substructure, and linkage disequilibrium across Europe. *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 continental sample 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_{EEU} is an indication of the number of other nuclear loci for which each focal nuclear locus exhibits a non-random association in the EEU deme; LD_{WEU} is an indication of the number of other nuclear loci for which each focal nuclear locus exhibits a non-random association in the WEU deme; and LD_{EU} is an indication of the number of other nuclear loci for which each focal nuclear locus exhibits a non-random association across the whole European sample, based upon an approximation of Fisher's exact test of random association and a *P*-value £ 0.05.

Locus	M	N	H_O	H_E	F_{IT}	LD_{EEU}	LD_{WEU}	LD_{EU}
SL_eSSR04	3	12	0.678	0.806	0.172	1	7	8
SL_eSSR06	3	13	0.516	0.818	0.37	5	8	9
SL_eSSR09	3	11	0.324	0.502	0.408	6	9	11
SL_eSSR12	2	15	0.656	0.876	0.294	6	6	11
SL_eSSR16	3	8	0.415	0.528	0.485	2	0	10
SL_eSSR20	3	5	0.051	0.258	0.819	7	9	11
SL_eSSR29	3	21	0.761	0.902	0.205	3	8	11
SL_eSSR30	3	18	0.362	0.797	0.544	3	6	10
slat_18	3	11	0.307	0.809	0.651	4	6	11
slat_32	4	6	0.42	0.789	0.487	4	3	7
slat_33	3	2	0.061	0.134	0.94	2	0	12
slat_48	3	2	0.11	0.25	0.562	3	4	5
slat_72	3	13	0.436	0.818	0.504	5	4	9
slat_85	3	14	0.2	0.536	0.77	3	10	13
Overall		***10.79***	***0.38***	***0.63***	***0.498***	***3.86***	***5.71***	***9.86***

Table 2. Values of Cyto-nuclear linkage disequilibrium (D or D') between the mitochondrial ATP1 SNP and the most common allele (at frequency p) of each of 14 nuclear microsatellite markers calculated from sample taken at two localities in Europe (EEU & WEU) and a mixture of those European samples (EU). Also presented are D and D" calculated from North American (NA) samples reported by Fields *et al.* (2014).

	EEU			*WEU*			*EU*			*NA*		
Locus	*P*	*D*	*D¢*	*P*	*D*	*D¢*	*P*	*D*	*D¢*	*P*	*D*	*D¢*
SL_eSSR04	0.329	0	0	0.237	-0.0288	-0.25	0.276	-0.0099	-0.0682	0.498	0.0255	0.1737*
SL_eSSR06	0.268	-0.0069	-0.061	0.395	0.0481	0.25	0.342	0.0137	0.1146***	0.262	-0.0257	-0.1751**
SL_eSSR09	0.744	0.0234	0.2361	0.623	-0.1202	-0.641	0.672	-0.0579	-0.4504**	0.743	0.0312	0.2161**
SL_eSSR12	0.232	-0.0386	-0.3889	0.061	0.0337	1	0.133	0.0223	0.387***	0.315	0.0448	0.3243***
SL_eSSR16	0.207	-0.022	-0.2254	0.816	0.0048	0.0526	0.561	-0.0513	-0.3266***	0.544	0.0539	0.3258***
SL_eSSR20	0.902	-0.011	-1	0.798	-0.0481	-0.5	0.842	-0.0237	-0.4466***	0.953	-0.0078	-0.3721
SL_eSSR29	0.25	-0.0413	-0.4762	0.026	-0.0144	-1	0.119	-0.0064	-0.0836***	0.203	0.0254	0.2841**
SL_eSSR30	0.427	0.0317	0.4107	0.298	-0.024	-0.1515	0.352	0.0057	0.042	0.389	0.0117	0.0691
slat_18	0.038	-0.022	-0.5926	0.429	-0.0481	-0.25	0.268	-0.0636	-0.4032***	0.402	0.0125	0.071
slat_32	0.351	0.0069	0.1316	0.278	0.0288	0.2143	0.308	0.0217	0.2088	0.316	-0.0561	-0.3161***
slat_33	0.11	-0.0262	-0.3016	0.965	0	0	0.607	-0.0748	-0.5362***	0.607	0.1042	0.4868***
slat_48	0.808	0.0096	0.0833	0.9	0.0337	0.6364	0.862	0.0174	0.1981	0.582	0.0475	0.2415***
slat_72	0.325	0.0193	0.3889	0.219	0.0673	0.5833	0.263	0.0518	0.5573*	0.308	0.0026	0.0193
slat_85	0.878	0.0358	0.3714	0.145	-0.0577	-0.75	0.458	0.0322	0.1995***	0.516	-0.0256	-0.1492*