

Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. III. Interlocus associations

Donald M. Waller and
Susan E. Knight

Department of Botany, University of Wisconsin,
430 Lincoln Drive, Madison, WI 53706, U.S.A.

Electrophoresis was used to examine levels of association between alleles at different variable loci within eight natural populations of the annual plant, *Impatiens capensis*. This species produces both obligately self-fertilizing cleistogamous (CL) and strongly protandrous chasmogamous (CH) flowers. Values of gametic (D_{ij}) and composite (Δ_{ij}) disequilibrium estimated for all pairs of covarying loci revealed significant levels of disequilibrium in most of the populations. A bootstrap technique was used to obtain weighted means and standard errors of normalized disequilibrium estimates and these were compared among populations and groups. Mean composite disequilibrium (Δ') spanned a remarkable range among populations (0.09–1.0, mean = 0.43), as did fixation indexes (f), but these were only weakly correlated with each other ($r = 0.50$). Levels of disequilibrium in the CL and CH progeny generally resembled those found in their maternal parents, but values for the CH group were more variable. A reduction in disequilibrium was sometimes associated with outcrossing, but it is unlikely that selection to reduce disequilibrium favours outcrossing in this species.

INTRODUCTION

The presence of association between alleles at different loci, usually termed gametic or linkage disequilibrium, can result from selection, recent mutations, the mixing of genetically distinct populations, or random drift in allele frequencies (Crow and Kimura 1970; Lewontin, 1974; Hedrick *et al.*, 1978). Consistent patterns of disequilibrium have often been interpreted as evidence for natural selection (Lewontin, 1974; Allard *et al.*, 1972*a, b*). Less regular patterns of association among alleles are likely to occur in small populations or under other circumstances where recombination is restricted (the "Hill-Robertson" effect, after Hill and Robertson, 1966). Plant populations often experience self-fertilization and/or local pollen and seed movement (Jain, 1976; Levin, 1981) and so might be expected to frequently exhibit significant levels of disequilibrium. Unlike inbreeding which can be eliminated in a single episode of random outcrossing, inter-locus associations may persist for several generations, even in the absence of chromosomal linkage between the loci.

Interlocus associations, if frequent, could play an important role in the evolution of mating sys-

tems by interfering with the fixation of new favourable alleles and/or the elimination of deleterious alleles (Felsenstein, 1974; Felsenstein and Yokoyama, 1976; Strobeck *et al.*, 1976; Maynard Smith, 1978; Crow, 1988). For example, deleterious alleles due to mutation are expected to accumulate by drift in small populations and could come to be associated with other favourable and deleterious alleles even in the absence of linkage. Under such circumstances, free sexual recombination could be favoured for its ability to reduce this randomly-generated disequilibrium and so facilitate simultaneous, independent selection at the different loci. These authors compare sexual and asexual populations, but the arguments could also provide an advantage for increased outcrossing in a selfing population (cf. Heller and Maynard Smith, 1978). That is, an allele that increases outcrossing could increase if it effectively dissipated levels of gametic disequilibrium that interfere with selection at other loci. This contrasts with the selective advantage most commonly invoked to explain the advantage of outcrossing over self-fertilization, namely classical inbreeding depression resulting from the increased expression of deleterious recessive alleles and/or reduced overdominance

in the more homozygous inbred progeny (Crow and Kimura, 1970; Falconer, 1981). Selfing resembles asexuality in transmitting a higher fraction of maternal genes and slowing the dissipation of disequilibrium. A relative excess of multiple homozygotes and multiple heterozygotes ("identity disequilibrium") is also expected in self-fertilizing populations, even in the presence of gametic equilibrium (Cockerham and Weir, 1968; Weir and Cockerham, 1973). Such identity disequilibrium plays a role in mating system evolution by affecting the likelihood of associative overdominance (Uyenoyama, 1988).

For outcrossing to be favoured over selfing for its ability to dissipate disequilibrium between alleles at separate loci, several conditions should be satisfied: levels of disequilibrium should be relatively large, the disequilibrium should be of the type that interferes with selection (rather than being the product of a consistent pattern of selection), and outcrossing should be effective in reducing disequilibrium, relative to self-fertilization.

Here we examine levels of gametic disequilibrium in several natural populations of the common woodland annual, *Impatiens capensis* to test whether these conditions are satisfied. This species mixes self- and cross-fertilization by producing two distinct types of flower: tiny, obligately self-fertilizing cleistogamous (CL) flowers, and showy, protandrous chasmogamous (CH) flowers commonly pollinated by bumblebees. We compare patterns of disequilibrium among populations and among maternal parents and the two progeny

types. Seedlings produced by the CH flowers frequently display a fitness advantage over selfed siblings (Waller, 1984; Mitchell-Olds and Waller, 1985) but the genetic basis for this advantage remains unclear. The CH flowers are mechanically incapable of self-fertilization (Wood, 1975), but pollen is often transferred between flowers on the same plant, resulting in geitonogamous selfing. The genetic and mating structure of these populations is reported elsewhere (Knight and Waller, 1987; Waller and Knight, 1989). For other information on the breeding system of *Impatiens*, see Schemske (1978) and Waller (1979; 1980).

METHODS

Electrophoretic survey

The study populations occupy five sites in southern Wisconsin and three sites near Princeton, New Jersey, U.S.A. (table 1). Sixty to 70 maternal plants were sampled from seven of these populations and 16 from the SF-N.J. population. The maternal plants were chosen by proximity to random points within a 2 m × 5 m grid at five of the sites. At the MZ, SF and PW sites, the plants were sparser and so were collected over a somewhat larger area. From each maternal plant, we collected at least four seeds from capsules known to be derived from CL flowers. Only larger plants growing under favourable circumstances tend to produce CH flowers, however. If mature CH capsules were

Table 1 Allele frequencies observed at variable loci in each population. Sample sizes refer to the number of CL and (CH) progeny analyzed from each population

Population (State)	Sample sizes	Allele	Variable loci						
			<i>Adh-2</i>	<i>Idh-1</i>	<i>Idh-2</i>	<i>Mdh</i>	<i>Fea</i>	<i>Gpt-1</i>	<i>Pgi-2</i>
PU (NJ)	208 (20)	1		0.644	0.865				0.663
		2		0.356	0.135				0.337
PW (NJ)	252 (0)	1			0.338				0.972
		2			0.662				0.028
MV (WI)	240 (40)	1			0.292		0.808	0.817	
		2			0.708		0.192	0.183	
ED (WI)	216 (28)	1		0.345	0.945	0.182	0.991	0.964	
		2		0.665	0.055	0.727	0.009	0.036	
		3				0.091			
GI (WI)	244 (96)	1		0.032	0.153			0.306	
		2		0.968	0.847			0.694	
MZ (WI)	214 (148)	1	0.976	0.016	0.226		0.847	0.806	0.992
		2	0.024	0.984	0.774		0.153	0.194	0.008
SF (NJ)	64 (64)	1		0.375	0.469			0.906	0.844
		2		0.625	0.531			0.094	0.156
AT (WI)	240 (128)	1			0.933	0.933		0.883	
		2			0.067	0.067		0.117	

present, we also collected four CH seeds (these can be distinguished from the CL structures—Schemske, 1978; Waller, 1980). Sample sizes are listed in table 1.

Gels and stains were prepared using conventional techniques (O'Malley *et al.*, 1980; Guries and Ledig, 1982). Of the 31 loci initially screened, eight loci encoding six enzymes proved to be polymorphic in one or more populations (Knight and Waller, 1987) and seven were polymorphic enough to use in this study (table 1). The linkage (chromosomal) relationships of these variable loci is unknown. The four necessarily selfed CL seeds from each plant provide eight gametes for inferring the maternal genotype with high reliability (the probability of miss-scoring a heterozygote = $(0.5)^7 = 0.0078$). Too few CH progeny were present in populations PW and PU to estimate disequilibria for this group. For further details about the sites, collections, and methods, see Knight and Waller (1987).

Estimating linkage disequilibrium

The commonest measure of association between alleles at distinct loci is gametic (or "linkage") disequilibrium. This quantity is estimated from the departure of observed gametic frequency from that expected by random association of alleles at the two loci:

$$D_{ij} = P^{ij} - p_i p_j$$

where allele i occurs at frequency p_i at locus 1, allele j at frequency p_j at locus 2, and they co-occur in gametes at frequency P^{ij} . When data on diploid genotypes are available instead of gametic data, estimating gametic disequilibrium requires making assumptions regarding population structure (Hedrick *et al.*, 1978; Weir and Cockerham, 1979). For example, one may estimate gametic frequencies from the diploid genotypes under the assumption that the population is panmictic. It is clearly inappropriate, however, to assume panmixia in a self-compatible species with non-random patterns of mating like *Impatiens*. Another alternative occurs when self-fertilization is frequent: double heterozygotes occur so rarely that it is possible to infer gametic types directly from the diploid genotypes (e.g., Brown *et al.*, 1977). Double heterozygotes must be excluded from such analyses since coupling and repulsion pairing of alleles can not be distinguished. We estimated gametic disequilibrium using this approach, but found an

appreciable frequency of double heterozygotes in several populations (range: 0–36.4 per cent).

To avoid assuming panmixia or a uniformly high selfing rate, we also followed the recommendation of Weir (1979) and Weir and Cockerham (1979) to estimate Burrow's composite measure of total gametic and non-gametic digenic disequilibrium:

$$\Delta_{ij} = (P^{ij} + P_j^i - 2 p_i p_j) = D_{wij} + 2 D_{bij}$$

where P_j^i is the estimated frequency of alleles i and j co-occurring in repulsion and D_{wij} represents the within-individual and D_{bij} the between-individual component of disequilibrium. To correct for bias, both D_{ij} and Δ_{ij} were adjusted by the factor $N/(N-1)$. With random mating, D_b is zero and $\hat{\Delta}_{ij}$ provides an unbiased estimate of D_{ij} . Under the hypothesis that delta is zero, the variance of its estimate is (Weir, 1979):

$$\text{Var } \Delta_{ij} = [p_i(1-p_i) + D_i][q_j(1-q_j) + D_j] / N$$

where D_i , the one-locus disequilibrium coefficient, is given by:

$$D_i = P_i - p_i^2.$$

We wrote a program to provide estimates of gametic disequilibrium, composite disequilibrium, and associated chi-square statistics for each pair of alleles, correcting for departures from random mating (Hill, 1974; Weir, 1979; Laurie-Ahlberg and Weir, 1979; Black and Krafur, 1985). Alleles with a frequency of less than 0.02 were excluded from analysis to avoid obtaining low expected frequencies of certain genotypes and inflated chi-square values. Gametic disequilibrium was estimated under the assumption that each of the N diploid genotypes represent an independently sampled gamete. Disequilibria were estimated separately for the maternal parents and their CL and CH progeny, making three groups. Significance levels for the individual disequilibria are reported only for the parental populations as the presence of common maternal parents violates the assumption of independence in the chi-square tests.

The magnitudes of both measures of disequilibrium depend on allele frequencies and range from zero (random association) to a maximum possible absolute magnitude of 0.25. To permit comparisons among loci with diverse allele frequencies and among populations, we also calculated Lewontin's (1964) normalized measure of gametic disequilibrium:

$$D'_{ij} = D_{ij} / D_{\max}$$

and, by analogy, a normalized measure of composite disequilibrium. Here, D_{\max} is the minimum of $[p_i q_j, (1-p_i)(1-q_j)]$ when D_{ij} is negative or the minimum of $[p_i(1-q_j), (1-p_i)q_j]$ when D_{ij} is positive. These normalized measures were then averaged over alleles and loci. For each particular pair of loci, k and l , mean gametic and composite disequilibria were estimated by taking a weighted average of the absolute values of the estimates over all pairs of alleles, i at k and j at l . For gametic disequilibrium this estimate is:

$$\hat{D}' = \frac{\sum_i \sum_j [p_i(1-p_i) + D_i][q_j(1-q_j) + D_j] |D'_{ij}|}{\sum_i \sum_j [p_i(1-p_i) + D_i][q_j(1-q_j) + D_j]}$$

(cf. Clegg *et al.*, 1972; Weir and Cockerham, 1979). Hedrick (1987) presents a similar formula that uses only the products of gene frequencies as weights. The weights used here represent the expected variances of the disequilibrium values and are therefore appropriate for inbred populations expected to exhibit non-random associations (B. Weir, personal communication). As only two alleles were expressed at most loci, these weighted averages were generally identical in magnitude to the originally estimated disequilibria. These means were then combined over all covarying pairs of loci in a population to estimate an overall mean normalized disequilibrium. The same weights as above were used, accumulated over all pairs of loci:

$$\text{estimated mean } D' = \frac{\sum_{k,l} \text{wt}_{k,l} D'_{k,l}}{\sum_{k,l} \text{wt}_{k,l}}$$

with an analogous formula for Δ . Pairs of loci with intermediate allele frequencies receive larger weightings by this procedure. This overall mean allows us to compare the parents and CL and CH progeny in each population. It implicitly assumes that each pair of loci provides an independent estimate of the mean disequilibrium.

We estimated approximate error statistics for the normalized disequilibrium estimates using a bootstrap procedure (Efron, 1982). Families were repeatedly randomly sampled from the original data set (with replacement) until the number of families equalled that in the original sample. This sampling procedure was then repeated a large number of times to obtain the series of new estimates, D'_n^* and Δ'_n^* . The bootstrapped estimate of

the standard deviation of the estimator D' is then:

$$S.D._{\text{boot}}(D') = [\text{Var}(D'^*)]^{1/2}$$

with an analogous formula for Δ' . Sampling the data 100 times has been found adequate in Monte Carlo studies (Efron, 1982, p. 28); we used 200–400 repeated samplings.

RESULTS

Levels of association among loci vary considerably among populations and among pairs of loci (table 2). Estimated mean normalized composite disequilibria (Δ') ranged from 0.094 in population PU to 1.0 in population PW only 2 km away. These two populations were also the most and least inbred, respectively (table 3), suggesting a correspondence between inbreeding and the extent of disequilibrium. Across all populations, however, estimated mean disequilibrium and population fixation index are only weakly correlated (fig. 1; $r = 0.502$, ns; note: 14 df would be necessary for this correlation to be significant at $P < 0.05$). Five of the eight populations exhibited levels of disequilibrium significantly ($P < 0.05$) greater than zero at at least one covarying pair of loci and 14 of the 32 individual values were significant. Because no statistical corrections were made for multiple comparisons, one or two of the only nominally significant disequilibria could occur by chance. Still, the number and strength of the associations in many populations suggest that disequili-

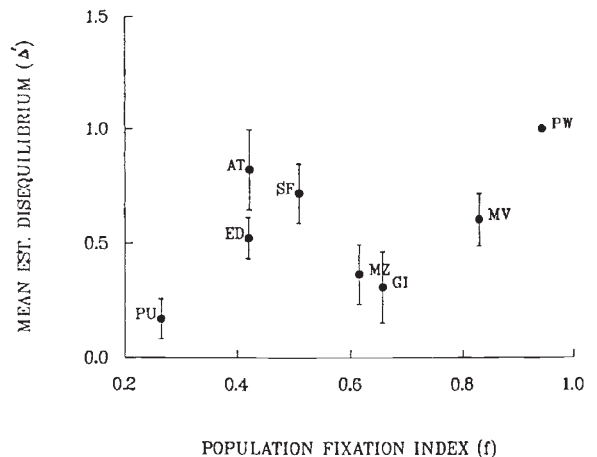


Figure 1 Scatterplot between estimated population disequilibria and estimated population fixation index (f). Estimates of mean normalized composite disequilibrium (Δ') are shown. Correlation with f is $r = 0.502$.

Table 2 Estimated mean gametic (D) and composite (Δ) disequilibria between pairs of variable loci. Only results for parental populations are shown. Normalized values (D' and Δ') are also shown

Population (State)	Variable loci	Estimated Disequilibria				Mean over loci
		D	D'	Δ	Δ'	Δ'
PU (NJ)	<i>Idh-1, Idh-2</i>	0.0492	0.0464	0.0004	0.002	0.094
	<i>Idh-1, Pgi-2</i>	-0.0335	-0.259	-0.0285	-0.117	
PW (NJ)	<i>Idh-2, Pgi-2</i>	-0.0320	-0.782	-0.0140	-0.151	1.00
	<i>Idh-2, Fea</i>	0.0098	1.00	0.0193*	1.00	
MV (WI)	<i>Idh-2, Gpt-1</i>	0.0130	0.226	0.0290	0.255	0.586
	<i>Idh-2, Gpt-1</i>	0.0461	0.840	0.0833***	0.766	
ED (WI)	<i>Fea, Gpt-1</i>	-0.0232	-0.711	-0.0545**	-0.763	0.508
	<i>Idh-1, Idh-2</i>	-0.0109	-1.00	-0.0264	-0.363	
GI (WI)	<i>Idh-1, Mdh</i>	0.0847	0.632	0.0958***	0.375	0.180
	<i>Idh-1, Gpt-1</i>	-0.0437	-1.00	-0.0485***	-1.00	
MZ (WI)	<i>Idh-2, Mdh</i>	-0.0046	-0.510	-0.0177*	-0.421	0.123
	<i>Idh-2, Gpt-1</i>	-0.0008	-1.00	-0.004	-1.00	
SF (NJ)	<i>Mdh, Gpt-1</i>	-0.0291	-0.693	-0.0387***	-0.676	0.697
	<i>Idh-1, Idh-2</i>	0.0114	0.410	0.0227**	0.410	
AT (WI)	<i>Idh-1, Gpt-1</i>	0.0093	0.527	0.0291**	0.640	0.800
	<i>Idh-2, Gpt-1</i>	0.0178	0.165	0.0111	0.051	
MV (WI)	<i>Adh-2, Idh-2</i>	0.0056	1.00	0.0029	0.262	0.123
	<i>Adh-2, Fea</i>	0.0044	0.213	0.0007	0.016	
SF (NJ)	<i>Adh-2, Gpt-1</i>	-0.0033	-1.00	-0.0095	-1.00	0.697
	<i>Idh-2, Fea</i>	0.0049	0.147	0.0048	0.068	
AT (WI)	<i>Idh-2, Gpt-1</i>	-0.0039	-0.0	-0.0095	-0.031	0.697
	<i>Fea, Gpt-1</i>	0.0273	0.2	0.0299	0.119	
SF (NJ)	<i>Idh-1, Idh-2</i>	0.1458	0.686	0.1583**	0.373	0.697
	<i>Idh-1, Gpt-1</i>	0.0455	1.00	0.0750*	1.00	
AT (WI)	<i>Idh-1, Pgi-2</i>	0.0758	1.00	0.1250**	1.00	0.800
	<i>Idh-2, Gpt-1</i>	-0.0284	-0.385	-0.0729*	-0.686	
AT (WI)	<i>Idh-2, Pgi-2</i>	0.0636	0.560	0.0896*	0.573	0.800
	<i>Gpt-1, Pgi-2</i>	-0.0119	-1.00	-0.0313	-1.00	
AT (WI)	<i>Idh-2, Mdh</i>	-0.0045	-1.00	-0.0090	-1.00	0.800
	<i>Idh-2, Gpt-1</i>	-0.0083	-1.00	-0.0073	-0.464	
AT (WI)	<i>Mdh, Gpt-1</i>	-0.0024	-0.309	-0.0158	-1.00	0.800

Nominal levels of significance, based on chi-squared tests with 1 df:

* $P < 0.05$.

** $P < 0.01$.

*** $P < 0.001$.

Table 3 Estimated mean adjusted composite disequilibria ($\hat{\Delta}'$) for parents and the two progeny groups. Estimated population fixation indices (f) are also shown. Mean disequilibria and standard errors are estimated via a bootstrap procedure

Population	f	Parents		CL Progeny		CH Progeny	
			(SE)		(SE)		(SE)
PW	0.943	1.000	(0.000)	1.000	(0.000)		
AT	0.422	0.822	(0.176)	0.806	(0.185)	0.605	(0.237)
SF	0.509	0.718	(0.130)	0.778	(0.135)	0.476	(0.169)
MV	0.830	0.603	(0.114)	0.595	(0.112)	0.827	(0.117)
ED	0.420	0.523	(0.090)	0.530	(0.084)	0.445	(0.348)
GI	0.656	0.308	(0.156)	0.375	(0.191)	0.268	(0.126)
MZ	0.615	0.364	(0.130)	0.350	(0.136)	0.242	(0.081)
PU	0.264	0.169	(0.087)	0.204	(0.104)	0.317	(0.097)
Averages:		0.501		0.520		0.454	

brum is of common occurrence in populations of this self-fertile annual.

Individual estimates of disequilibria vary widely among pairs of loci within populations (table 2), as do their normalized values. Generally, estimated values of gametic and composite disequilibrium resemble each other although differences are also apparent in some cases (table 2). Among the overall population means, normalized estimates of gametic disequilibrium (not shown) were generally greater than, but highly correlated with, estimates of normalized composite disequilibrium ($r=0.918$). Such might be expected if the estimates of gametic disequilibrium were biased upward by ignoring all double heterozygotes. For this reason, only estimated composite disequilibria are discussed further.

The bootstrap estimates of mean normalized composite disequilibrium were generally close to, but occasionally higher than, the estimates obtained directly from the data (cf. tables 2 and 3). The differences are greatest for the three populations with the lowest estimated disequilibrium (GI, MZ and PU) where the bootstrap estimates are up to two or three times greater than the direct estimates. The estimated standard errors suggest that all the populations have mean composite disequilibrium greater than zero except perhaps PU (parental values—table 3). Among the progeny, only the estimated mean disequilibrium (0.445) for CH progeny in population ED is not demonstrably significantly greater than zero, and this appears attributable to the large estimated standard error (0.348) associated with the small CH sample size (table 1).

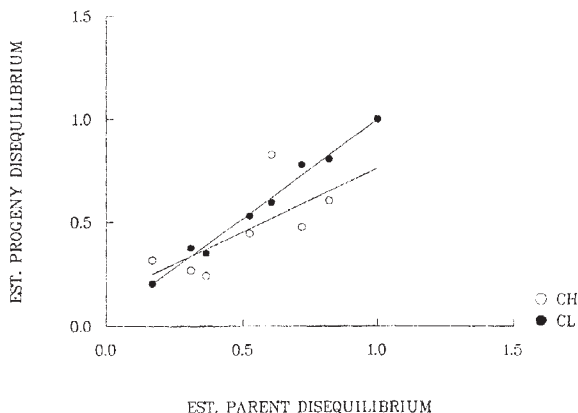


Figure 2 Scatterplot of mean adjusted composite disequilibrium (Δ) for the two progeny groups (CL—filled circles, and CH—empty circles) as a function of parental disequilibrium.

Estimated mean disequilibria in the progeny groups generally mirrored levels of disequilibrium found in their maternal parents (table 3 and fig. 2). This correspondence is particularly close between CL progeny and their parents ($r=0.993$) and somewhat less between CH progeny and parents ($r=0.690$). Five of seven populations exhibited lower CH than parental disequilibrium, and this reduction ranged up to 34 per cent (populations SF and MZ). None of these differences was statistically significant, however, and populations PU and MV showed apparent increases in disequilibrium. These results suggest that while outcrossing can be effective in breaking up particular maternal allelic combinations, other associations can also be introduced via the paternal genome.

DISCUSSION

Alleles at distinct loci often tend to be associated with each other in natural populations of *I. capensis*. Such associations are not unexpected in a self-fertilizing species with an inbred population structure like *Impatiens*. Almost half of the 2-locus disequilibria deviated significantly from zero, with five of the eight populations containing at least one pair of alleles with significant association. The linkage relations among these loci are unknown, but the relatively high number of chromosomes ($2N=20$; Wood, 1975) make it unlikely that the few variable loci examined in any one population are contained in a tight linkage group. Interestingly, population means of normalized disequilibria over all loci (mean $\hat{\Delta}$) exhibited even greater levels of apparent statistical significance based on the bootstrap results. All the populations (except perhaps PU) had disequilibria significantly greater than zero, with estimated standard errors that ranged from 0.00 to 0.176. The magnitude of these estimated errors do not appear to depend directly on sample size in these populations (e.g., the two largest values, 0.237 and 0.348 for CH progeny in AT and ED, occurred in populations with the second largest and second smallest sample sizes).

Linkage disequilibrium can be generated both by random drift of allele frequencies in small and/or inbred populations and by selection favouring particular genetic combinations at different loci. With population subdivision, alleles at separate loci can also come to be correlated with each other, even with linkage equilibrium within the subpopulations (Nei and Li, 1973; Ohta, 1982). Thus, to identify selectively based disequilibrium, it is necessary to demonstrate the presence of con-

sistent linkage disequilibrium across several populations (Lewontin, 1974). We did not find disequilibrium of consistent sign or magnitude between particular pairs of loci among the populations we examined (table 2). This suggests that much of the observed disequilibrium may be randomly generated or due to inconsistent selection at these or linked loci.

Significant levels of randomly generated disequilibrium are to be expected in populations of species like *I. capensis* due to frequent inbreeding, founder effects, and consequent small effective population sizes. A weak correlation was observed between estimated mean disequilibrium and population fixation index ($r = 0.50$), suggesting that factors that promote inbreeding may also help maintain disequilibrium. Other factors, such as historical patterns of colonization and population subdivision are probably just as important, however. The considerable variation found among pairs of loci within populations and among populations is consistent with the hypothesis that this disequilibrium results more from random events than a consistent pattern of selection. Alternatively, variation among the populations could also be due to chromosome aberrations restricting recombination in certain populations.

Outcrossed animal populations usually exhibit little disequilibrium except between tightly linked loci (e.g., *Drosophila montana*—Baker, 1975; *D. melanogaster*—Langley *et al.*, 1977). Two of ten pairs of loci displayed significant disequilibrium in a natural house fly population (Black and Kraf-sur, 1985). These involved rare alleles and were not consistent across populations and so also presumably reflect random population processes. Similar, 32 of 50 Yanomama villages in South America exhibited significant disequilibrium (mostly from two pairs of closely linked loci) that appeared to result from population subdivision and inbreeding rather than any consistent pattern of selection (Smouse and Neel, 1977). In a rare study of a self-fertile animal species, Foltz *et al.* (1982) documented levels of disequilibrium in eight populations of the self-fertile slug, *Deroceras laeve* that were considerably greater than those found in an outcrossed relative. Again, the disequilibrium was not consistent and was therefore interpreted to be the result of drift.

In contrast to these results, significant levels of disequilibrium have often been observed in self-fertile plants where it has frequently been interpreted as evidence for consistent selection. For example, two California populations of wild oats (*Avena barbata*) exhibited marked disequilibrium

between both linked and unlinked loci that appeared to be related to the plants' microenvironment (Allard *et al.*, 1972*b*). Similarly, disequilibrium consistently increased over many generations of self-fertilization following the creation of composite crosses in barley (*Hordeum vulgare*), apparently due to selection and restricted recombination (Clegg *et al.*, 1972; Weir *et al.*, 1972). Such patterns are not universal, however. Eleven mostly selfing Israeli populations of wild barley (*H. spontaneum*) also displayed significant disequilibrium, but not in any obvious ecological pattern (Brown *et al.*, 1977). Natural populations of *Lupinus succulentus* displayed no significant linkage disequilibrium even between a pair of linked loci (Harding and Barnes, 1977). This last species most resembles *Impatiens* in being variably outcrossed, insect (honeybee) pollinated, and a colonizer.

For outcrossing to be favoured over selfing for its ability to reduce disequilibrium, we reasoned that three conditions should be satisfied (see Introduction). The patterns of disequilibrium described here appear to satisfy the first two, namely that significant levels of disequilibrium exist in many populations, and that this disequilibrium appears to be of the kind that might interfere with simultaneous selection at several loci (*i.e.* it appears to be generated by a variety of extrinsic forces rather than reflecting a consistent pattern of selection). However, the third condition (that outcrossing be effective in reducing disequilibrium) has not been demonstrated. Generally, both CL and CH progeny display levels of disequilibrium similar to their maternal parents (fig. 2). Although levels of disequilibrium among the CH progeny are more variable, they were neither consistently nor significantly lower than their parents. This may not be surprising, however, given the single elapsed generation and the genetic consequences of CH reproduction in this species. Although less inbred than CL progeny, CH progeny experience both moderate levels of geitonogamous selfing (mean = 50 per cent) and frequent biparental inbreeding as the result of consanguineous matings (Waller and Knight, 1989).

The rate at which disequilibrium decays depends upon the rates of both recombination and outcrossing (Weir and Cockerham, 1972) and, in the case of subdivided populations, on the rates of migration (Nei and Li, 1973). Rates of migration are difficult to estimate but appear to be on the order of 0.3 to 3 per population per generation (Knight and Waller, 1987). The existence of two populations in which CH disequilibria were higher than parental values also demonstrates that out-

crossing may not always act to reduce disequilibrium. Such increases could result from pollination involving one or a few pollen parents with distinct genotypes.

Thus, the associations between alleles at distinct loci that occur relatively often in *I. capensis* are not uniformly or rapidly dissipated in a single generation of "outcrossing". Inbreeding and limited opportunities for gene exchange probably contribute to the high levels of disequilibrium in several populations. Nevertheless, estimated levels of disequilibrium are not uniformly high across all populations. This suggests that chasmogamy and its consequent outcrossing can, at least in some instances and over time, effectively dissipate disequilibrium. Such effects, however, do not appear frequent or strong enough in *I. capensis* to consistently favour outcrossing through short-term, individual selection.

Acknowledgements H. and E. Horn and J. Schmitt helped collect the seeds and D. O'Malley and R. Guries provided experienced guidance with the electrophoresis. J. Crow, T. Mitchell-Olds, P. Hedrick and B. Weir provided useful suggestions on an earlier version of the manuscript. The National Science Foundation (DEB 82-06996 and BSR-8421272) provided research support and a fellowship from the Alexander von Humboldt Foundation (Bonn, W. Germany) afforded time to fully analyze the results.

REFERENCES

- ALLARD, R. W., BABEL, G. R., CLEGG, M. T. AND KAHLER, A. L. 1972a. Evidence for coadaptation in *Avena barbata*. *Proc. Natl. Acad. Sci. (USA)*, **69**, 3043-3048.
- ALLARD, R. W., KAHLER, A. L. AND WEIR, B. S. 1972b. The effect of selection on esterase allozymes in a barley population. *Genetics*, **72**, 489-503.
- BAKER, W. K. 1975. Linkage disequilibrium over space and time in natural populations of *Drosophila montana*. *Proc. Natl. Acad. Sci. USA*, **72**, 4095-4099.
- BLACK, W. C. AND KRAFSUR, E. S. 1985. A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theor. Appl. Genet.*, **70**, 491-496.
- BROWN, A., NEVO, E. AND ZOHARY, D. 1977. Association of alleles at esterase loci in wild barley *Hordeum spontaneum*. *Nature*, **268**, 430-431.
- CLEGG, M. T., ALLARD, R. W. AND KAHLER, A. L. 1972. Is the gene the unit of selection? Evidence from two experimental plant populations. *Proc. Natl. Acad. Sci. USA*, **69**, 2474-78.
- COCKERHAM, C. C. AND WEIR, B. S. 1968. Sib mating with two linked loci. *Genetics*, **60**, 629-640.
- COCKERHAM, C. C. AND WEIR, B. S. 1977. Digenic descent measures for finite populations. *Genet. Res. (Cambridge)*, **30**, 121-147.
- CROW, J. F. 1988. The importance of recombination. In Michod, R. E. and Levin, B. R. (eds) *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Inc., Sunderland, MA, pp. 56-73.
- CROW, J. F. AND KIMURA, M. 1970. *An Introduction to Population Genetics Theory*. Burgess Publ. Co., Minneapolis.
- EFRON, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- FALCONER, D. S. 1981. *Introduction to Quantitative Genetics* (2nd ed.). Longman, New York.
- FELSENSTEIN, J. 1974. The evolutionary advantage of recombination. *Genetics*, **78**, 737-756.
- FELSENSTEIN, J. AND YOKOYAMA, S. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics*, **83**, 845-859.
- FOLTZ, D. W., SCHAITKIN, B. M. AND SELANDER, R. K. 1982. Gametic disequilibrium in the self-fertilizing slug *Deroceras laeve*. *Evolution*, **36**, 80-85.
- GURIES, R. P. AND LEDIG, F. T. 1982. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution*, **36**, 387-402.
- HARDING, J. AND BARNES, K. 1977. Genetics of *Lupinus*. X. Genetic variability, heterozygosity and outcrossing in colonial populations of *Lupinus succulentus*. *Evolution*, **31**, 247-255.
- HEDRICK, P. W. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics*, **117**, 331-341.
- HEDRICK, P. W., JAIN, S. AND HOLDEN, L. 1978. Multilocus systems in evolution. *Evol. Biol.*, **11**, 101-182.
- HELLER, R. AND MAYNARD SMITH, J. 1978. Does Muller's ratchet work with selfing? *Genet. Res. Camb.*, **32**, 289-293.
- HILL, W. G. AND ROBERTSON, A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res., Camb.*, **8**, 269-294.
- HILL, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, **33**, 229-239.
- JAIN, S. K. 1976. The evolution of inbreeding in plants. *Ann. Rev. Ecol. Syst.*, **7**, 469-495.
- KNIGHT, S. E. AND WALLER, D. M. 1987. The genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. I. Population-genetic structure. *Evolution*, **41**, 969-978.
- LANGLEY, C. H., ITO, K. AND VOELKER, R. A. 1977. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. *Genetics*, **86**, 447-454.
- LAURIE-AHLBERG, C. C. AND WEIR, B. S. 1979. Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. *Genetics*, **92**, 1295-1314.
- LEVIN, D. A. 1981. Dispersal versus gene flow in plants. *Annals Mo. Bot. Garden*, **68**, 232-253.
- LEWONTIN, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49-67.
- LEWONTIN, R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia Univ. Press, New York.
- MAYNARD SMITH, J. 1978. *The Evolution of Sex*. Cambridge University Press, Cambridge.
- MITCHELL-OLDS, S. T. AND WALLER, D. M. 1985. Relative performance of selfed and outcrossed progeny in *Impatiens capensis*. *Evolution*, **39**, 533-544.
- NEI, M. AND LI, W. H. 1973. Linkage disequilibrium in subdivided populations. *Genetics*, **75**, 213-219.
- OHTA, T. 1982. Linkage disequilibrium with the island model. *Genetics*, **101**, 139-155.
- O'MALLEY, D., WHEELER, N. C. AND GURIES, R. P. 1980. A manual for starch gel electrophoresis. Staff Paper, College of Agriculture and Life Sciences, Univ. of Wisconsin, Madison.

- SCHEMSKE, D. S. 1978. Evolution of reproductive characteristics in *Impatiens* (Balsaminaceae): The significance of cleistogamy and chasmogamy. *Ecology*, 59, 596-613.
- SMOUSE, P. E. AND NEEL, J. V. 1977. Multivariate analysis of gametic disequilibrium in the Yanomama. *Genetics*, 85, 733-752.
- STROBECK, C., MAYNARD SMITH, J. AND CHARLESWORTH, B. 1976. The effects of hitchhiking on a gene for recombination. *Genetics*, 82, 547-588.
- UYENOYAMA, M. 1988. On the evolution of genetic incompatibility systems: Incompatibility as a mechanism for the regulation of outcrossing distance. In Michod, R. E. and Levin, B. R. (eds) *The Evolution of Sex: An Examination of Current Ideas*, Sinauer Associates, Inc., Sunderland, MA, pp. 212-232.
- WALLER, D. M. 1979. The relative costs of selfed and outcrossed seeds in *Impatiens capensis* (Balsaminaceae). *Amer. J. Bot.*, 66, 313-320.
- WALLER, D. M. 1980. Environmental determinants of outcrossing in *Impatiens capensis* (Balsaminaceae). *Evolution*, 34, 747-761.
- WALLER, D. M. 1984. Differences in fitness between seedlings derived from cleistogamous and chasmogamous flowers in *Impatiens capensis*. *Evolution*, 38, 427-440.
- WALLER, D. M. AND KNIGHT, S. 1989. Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*. II. Outcrossing rates and genetic correlations. *Evolution* (In press).
- WEIR, B. S. 1979. Inferences about linkage disequilibrium. *Biometrics*, 35, 235-254.
- WEIR, B. S., ALLARD, R. W. AND KAHLER, A. L. 1972. Analysis of complex allozyme polymorphisms in a barley population. *Genetics*, 72, 505-523.
- WEIR, B. S. AND COCKERHAM, C. C. 1973. Mixed self and random mating at two loci. *Genet. Res., Camb.*, 21, 247-262.
- WEIR, B. S. AND COCKERHAM, C. C. 1979. Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, 42, 105-111.
- WOODS, C. S. 1975. The Balsaminaceae in the Southeastern United States. *J. Arnold Arbor.*, 56, 413-426.