Population Biology of *Avena*

IV. Polymorphism in Small Populations of *Avena fatua**

S. K. JAIN and K. N. RAI

Department of Agronomy and Range Science, University of California, Davis (USA)

Summary. The population structure of wild oats (*Avena fatua*) sampled in two prune orchards was described using Wright's model of a population having many largely isolated, small subdivisions. A high degree of genetic differentiation was observed among the individual colonies for lemma color, leaf sheath hairiness and isoenzymatic loci. Estimates of genotypic frequencies and population sizes over a two-year period suggested that random drift played an important role in the population changes toward a highly mosaic pattern of differentiation and local monomorphism in a substantial proportion of colonies. It was recognized, however, that without additional extensive field studies, the hypothesis of irregularly dispersed factors of multiniche selection could not be ruled out. Similar studies are briefly reviewed in order to outline the research needed on the issue of selection versus random drift as the primary force in local differentiation.

Introduction

The widespread occurrence of genetic polymorphisms in *Avena fatua* populations of Central California was reported by Imam and Allard (1965) and Marshall and Jain (1968). Materials for these studies were derived invariably from large, more or less continuous roadside or range stands, with plant densities in the range of 200 to 2000 plants/meter². Rai and Jain (unpublished data), based on their studies of gene flow, obtained neighborhood size estimates in the range of 150 to 600. Selective forces were therefore estimated using a deterministic model ignoring random drift, which suggested heterozygote advantage as one of the factors maintaining polymorphisms. In one population near Marysville, clinal variation was analyzed in terms of selection-migration balance (Jain and Marshall, in preparation). However, recent studies of the population regulation and detailed spatial distribution of *Avena* over time suggest that on a very local scale, 'extinction' and 'recolonization' occurs intermittently resulting in drift through founder effects. Such events are rarely recorded in the usual surveys based on large continuous populations.

Avena fatua populations resident in many California orchards, however, offer an excellent opportunity for studying the genetic changes involving significant drift effects. Interculture operations during the winter and spring each year allow only small colonies to survive within 2 or 3 feet around the tree base, which reproduce successfully every summer. These colonies are small, largely isolated from each other in terms of gene flow, and easily sampled individually for census data and seed collection. Based on our

observations over the preceding years and known history, two prune orchards (referred to as 1 and 2), about 40 miles north of Davis, were selected having a notable difference in their relative colony sizes (due to the nature of tillage operation). Data on population size, gene flow and genetic variation are presented in relation to the evidence for the role of random drift in genetic differentiation and discussed in the context of a review of similar studies.

Materials and Methods

During late May of 1970 and 1971 a set of 50 trees in orchard 1 (smaller colony size) and of 40 trees in orchard 2, from eight adjacent rows within each orchard, were sampled by harvesting mature plants. Harvest time is adjusted to the stage of most seeds having shattered naturally so that sampling causes a minimum disturbance in relation to the seed input for the succeeding generation. Plant counts provided direct estimates of the actual colony sizes. Individual plants were scored for lemma color and leaf sheath pubescence (loci *b* and *ls*). Bulk seed was used for the estimation of heterozygosity and outcrossing rate using the socalled bulk method of Jain and Marshall (1968). Seed migration was recorded in terms of a census of seedlings around the colonies and in the intervening areas among trees. Effective migration is in fact lower due to the interculture operations removing those seedlings. Pollen and seed migration curves were also studied in several other areas (Rai and Jain, in preparation). A subsample of 16 colonies from orchard 1 was sampled in the Fall 1972 for isoenzymatic variation assays using starch gel electrophoresis.

Results

The plant census data are summarized in Table 1. Within each orchard and each year, there was a very wide range of plant numbers per colony and the arithmetic means show orchard 1 to have a much smaller mean colony size than orchard 2. In year 1971, the oat densities were higher than in 1970 due

to a good seed crop in 1970. Assuming that individual colony sizes have varied in time as much as their variation over space and two years indicates, the harmonic means would be more appropriate measure of effective sizes as suggested by Wright (1951, and earlier). Furthermore, since the mean seed output per plant generally has a high variance among plants, following Kimura and Crow (1963), the mean effective size N_e is given by

$$N_e = \frac{2N}{(1 - \alpha) + (1 + \alpha)\ V_k/\bar{k}},$$

where $\alpha = F$ = fixation index, $\bar{k}$ and V_k are mean and variance for fecundity. Measurements of seed output on a large number of wild oat populations suggest that $V_k/\bar{k}$ varies from 1 to as high as 5; here, we use $\alpha = F = .92$ (see Table 5), and $V_k/\bar{k} = 2$, so

Table 1. *Estimates of Mean Effective Colony Size*

Orchard	Year	$\bar{N}_a$	Range	$N = \bar{N}_h$	$\bar{N}_e = N/C$
1	1970	38	17 — 84	28.6	14.6
	1971	47	26—102	62.0	31.6
2	1970	70	21 — 176	47.9	24.4
	1971	123	48 — 191	79.0	40.3

$\bar{N}_a$ = Arithmetic mean; $\bar{N}_h$ = harmonic mean; $\bar{N}_e$ = effective size; K = variance/mean ratio for seed output/plant;

$$N_e = \frac{2N}{1 - \alpha + (1 + \alpha)\ V_k/\bar{k}};$$

α = measure of departure from random mating; so that

$$C = \frac{(1 - \alpha) + (1 + \alpha)\ V_k/\bar{k}}{2}.$$

that an adjustment constant for fecundity and inbreeding in $N_e = N/c$, is $c = 1.96$, on an average. Note from the last column in Table 1 that the effective colony size in both orchards are likely to be indeed much smaller than the actual numbers.

Evidence for almost complete isolation among colonies (trees are on a 18- to 19-foot square grid) comes from several sources. Independent detailed work on several populations indicated that both seed and pollen migration distances are small and that the distance curves might be leptokurtic (Rai and Jain, loc. cit.). On an average, 99% of the seed and pollen movement for any given generation interval is confined to within an area of 10-foot radius around the "source" or parental location. Seedling census in areas outside the colonies and intervening among the trees suggested that less than 1% seed survived and contributed to the rate of effective migration. Pollen flow is further restricted by the fact that *A. fatua* is a predominantly self-fertilizing species; estimates for the orchard populations were of the order of 94 to 97% selfing. Thus, these populations approach closely the model of a large population with almost completely isolated small subdivisions (Wright, 1951).

For loci *b* (black versus gray lemma color) and *ls* (hairy versus nonhairy leafsheath), individual colonies were scored for the proportion of dominant $(1 - R)$ and recessive (R) classes. Table 2 gives a representative sample of frequency data for a 5×5 matrix of trees for both loci and years. Besides the spatial distribution, frequency data are also summarized in the form of frequency distributions as given in Table 3. Note that orchard 1 distributions are *J*-shaped with a fairly high number of colonies in the fixed classes (gene frequency 0 or 1). Two points are to be noted about the comparison between the two years: (1) In some cases a colony may show fixation in the year 1970 but polymorphism in 1971. This is most likely due to the 10—20% seed carryover in the soil as estimated by direct counts of ungerminated seed, from one generation to the next. (2) The frequencies are highly correlated between the two years (Orchard 1: locus *b*, $r = .916$, locus *ls*, $r = .763$;

Table 2. *Representative Sample of 5 × 5 Matrix of Recessive Frequency (R)*
(Orchard 1)
Column

Row Year	Locus	1		2		3		4		5	
		1970	1971	1970	1971	1970	1971	1970	1971	1970	1971
1	bb	.971	.774	0.40	.048	.037	0	.374	.404	0	.004
	ls ls	1	1	1	1	.900	1	1	1	.970	1
2	bb	.053	.091	.360	.562	.068	.174	.021	0	0	.079
	ls ls	.458	.656	.286	.315	.920	1	1	1	1	.976
3	bb	.068	.011	.090	.072	.205	.136	.118	.057	×	
	ls ls	.286	0	.824	.978	.556	.844	.741	1	×	
4	bb	.076	.185	.118	.150	.843	.823	×		×	
	ls ls	.882	.90	.864	1	.875	.958	×		×	
5	bb	.381	.333	.049	.070	.021	0.56	0	.040	.034	.039
	ls ls	.950	1	1	1	.943	1	1	.962	1	1

× = tree missing

S. K. Jain and K. Rai: Population Biology of Avena. IV.

9

Table 3. *Frequency Distribution of Gray (bb) and Nonhairy Leaf Sheath (ls ls) Individuals*

| Orchard | Genotype | Year | Proportion of Subdivisions in the Frequency Class | | | | | | | No. Sample N |
			0.0	0—.199	.2—.399	.4—.599	.6—.799	.8—.999	1.0	
1	bb	1970	.326	.347	.102	.061	.020	.143	0	49
		1971	.220	.460	.060	.080	.100	.080	0	50
	ls ls	1970	0	0	.042	.042	.042	.426	.447	47
		1971	.018	0	.018	0	.056	.426	.481	54
2	bb	1970	.059	.352	.236	.235	.118	0	0	17*
		1971	.026	.410	.205	.205	.077	.052	.026	39
	ls ls	1970	.059	.059	.059	.118	.118	.470	.118	17*
		1971	.026	.026	.026	.026	.154	.436	.308	39

N = Number of colonies (individual trees) sampled.
* The remainder colonies could not be scored in 1970 as all the seed had shattered prior to scoring plants.

Orchard 2: locus b, $r = .861$, and locus ls, $r = .782$). All values of r are highly significant ($P < .01$).

The spatial patterns of intercolony differentiation was tested as follows: Since heterozygosity estimates could not be obtained from rather small seed samples available for individual colonies, phenotypic frequencies had to be used. Three statistical tests were carried out on the transformed values (φ) using the arc sine transformation $\varphi = \sin^{-1} \sqrt{R}$. (1) For the matrix in each orchard, column and row means were calculated and tested for a clinal (linear) gradient with a rank correlation. For example, if rows 1 to 8 have mean values of R_1, R_2, . . . , R_8, a regular cline would predict them to be in a rank order 1 to 8, which is matched against the observed rank order. These tests gave insignificant rank correlations ($P > > .05$) thus showing no evidence for clines along the directions of row or columns. (2) The correlation coefficients between neighborhoods taken pairwise for all colonies in orchard 1 and year 1970 were $r = — .084$ ($P > .05$) for locus b and $r = .245$ for locus ls ($P > .05$). (3) Finally, an overall heterogeneity was tested ignoring any spatial arrangement of colonies using χ^2 test gave highly significant χ^2 values ($P < .01$) in all cases. Thus, although there seems to be no spatial pattern in the distribution of genotypic frequencies, the colonies within each orchard are significantly differentiated in a more or less random or mosaic manner.

In order to verify the heterogeneity of genetic variation further, a sample of 16 colonies from orchard 1 was used for an isoenzymatic study using starch gel electrophoresis. A total of 24 different bands were scored for three systems (esterase, acid phosphatase and cathodal peroxidase); only five of them have been analyzed in terms of monogenic ratios. Here, we shall use only presence-absence data for the others to score a measure of the degree of polymorphism in terms of percent polymorphic "loci" per colony. The results are summarized in Table 4. Isozyme variation is measured in terms of the proportion of "loci" polymorphic within each colony whereas variation at loci b and ls is combined into

a phenomorphism index ($P I = 4 \, \Sigma \, R \, (1 — R)/n$, $n = 2$, the number of loci). Spearman's rank correlation between percent polymorphic loci for isozyme data and $P I$ based on loci b and ls gave a value of $r = .488$ ($P \sim .05$). Thus, isozyme variation seems to be weakly correlated with the random pattern observed for the two morphological markers. In several cases neighborhoods are likewise fixed for different "alleles" (bands). Clearly, more extensive genetic analysis and sampling would be of great interest.

Table 4. *Comparison of Variation at Isoenzymatic and Morphological Markers*

| Colony No. | Isoenzymatic % Polymorphic Loci* | Morphological Markers | | |
		R_{bb}	$R_{ls\,ls}$	$P\,I$
1	17	.053	.458	.597
2	23	.064	.920	.267
3	17	.021	1	.041
4	26	0	.833	.278
5	8	0	.875	.219
6	0	0	1	0
7	24	.090	.824	.227
8	21	.205	.556	.820
9	12	.118	.741	.592
10	0	.034	1	.066
11	15	.182	.947	.398
12	19	.076	.882	.349
13	21	.118	.864	.443
14	17	0	1	0
15	0	0	0	0
16	12	0	0	0

These limited data do not permit any further analysis of individual factors of evolution such as selection involving heterozygote advantage. We can, however, test at a gross level the change between the two years in relation to the predictions from a no-selection model: (1) That genotypic frequencies on an average are expected to show no change (i.e., $\overline{\Delta R}$ or $\overline{\Delta \varphi} = 0$), but (2) perhaps an increase in variance ($\overline{\sigma_\varphi^2}$) due to random dispersion; moreover, (3)

Tabelle 5. *Some Statistics of Genotypic Frequency Change*

Year Orchard	Locus	N	1970 $\bar{\varphi}$	1970 s_φ^2	1971 $\bar{\varphi}$	1971 s_φ^2	$\overline{\Delta\varphi}$	$\overline{|\Delta\varphi|}$	$s_{\Delta\varphi}^2$	t	F-ratio
1	b	49	23.42	607.17	23.47	720.64	0.05	7.7	116.04	0.04	1.19
	$1s$	47	74.19	361.62	76.97	421.22	2.78	10.2	187.21	1.41	1.16
2	b	17	32.10	299.99	30.44	271.67	−1.66	6.8	79.86	0.77	0.90
	$1s$	17	60.32	588.35	68.49	595.72	8.18	10.7	258.87	2.10*	1.01

N = Number of colonies scored.

φ = $\sin^{-1}\sqrt{R}$, R = % bb or $1s1s$ plants.

t = $\overline{\Delta\varphi}/s_{\Delta\varphi}$; $F = \dfrac{s_\varphi^2 \,(1971)}{s_\varphi^2 \,(1970)}$

$\overline{|\Delta\varphi|}$ Measures the change in φ, ignoring sign; compare this with $s_{\Delta\varphi}^2$ — both measure random drift effects.

* . . . .

variance of $\Delta\varphi(\sigma_{\Delta\varphi}^2)$ should be the same for different loci but different between the two orchards in view of their different colony sizes. Accordingly, Table 5 summarizes some statistics of genotypic frequency changes. Note that $\overline{\Delta\varphi}$ is significantly different from zero only in the case of orchard 2, locus $1s$. Both $(\Delta\varphi)$ and $s_{\Delta\varphi}^2$ measure random drift effects, assuming no selection. For orchard 2, locus b gave smaller values than orchard 1 but opposite is the case for locus $1s$. Note that $s_{\Delta\varphi}^2$ values of the two years are not significantly different. Overall these data suggest with the exception of frequencies at locus $1s$ in orchard 2 that most frequency changes between 1970 and 1971 are due to random drift and/or sampling errors in survey. In order to calculate expected $\sigma_{\Delta\varphi}^2$ under the assumption of steady state, we use the following formula due to Wright (1951, and pers. communication). In a large population subdivided into numerous colonies of size N and given m = gene flow, s = proportion of selfing,

$$F = (1-m)^2 \left[\frac{s}{2} + \left(\frac{1}{2N} - \frac{s}{2} \right)(1-m) \right] \Big/$$

$$\left[1 - \left(1 + \frac{s}{2} - \frac{1}{N} \right)(1-m)^2 - \left(\frac{1}{2N} - \frac{s}{2} \right)(1-M)^4 \right]$$

and

$$\sigma_{\Delta\varphi}^2 = \bar{q}\,(1-\bar{q})\,(1+F)/2N \,.$$

Values of F are given in Table 6 for $N = 25$, 50 and 100, migration rate (island model) $m = .01, .05$, and $s = .85, .90, .95$ and $.99$. For our data, the relevant values of F are those for $s = .95$, $m = .01$ and

Table 6. *Expected Values of F_e*

s	$N = 25$ $M = .01$	$N = 25$ $.05$	$N = 50$ $.01$	$N = 50$ $.05$	$N = 100$ $.01$	$N = 100$ $.05$	Single Large Population
.85	.871	.672	.828	.650	.788	.638	.739
.90	.908	.719	.869	.713	.842	.695	.818
.95	.929	.769	.914	.761	.900	.756	.905
.99	.955	.811	.952	.810	.948	.809	.980

$N = 25$ or 50, i.e., $F = .929$ and $.914$, respectively. The expected value of $\varphi_{\Delta q}^2 = \dfrac{\bar{p}\,\bar{q}\,(1+F)}{2\,N}$, which gives for locus b: $.0083$ (orchard 1) and $.0125$ (orchard 2) match the corresponding observed values: $.0085$ and $.0105$. For locus b, we obtained estimates of heterozygosity $\bar{H} = .045$ (orchard 1) and $.068$ (orchard 2), using seed bulks from pooled colonies. Correcting for Wahlund effect of subdivision, we have $F = 1 - \dfrac{\bar{H}}{2\,\bar{p}\,\bar{q} - 2\,V_p}$, which gives $.691$ (orchard 1) and $.662$ (orchard 2).

This provides a measure of the rate of random dispersion at steady state. However, it is not certain whether orchard populations are in equilibrium; the age of these orchards is only 10 years so that prior to the planting of trees we must assume a similar history of differentiation. In any event, random dispersion appears to have played an important role in the arrival of genetic variation pattern as we see it now.

Discussion

There has been a great deal of discussion of the relative importance of selection and random drift, as witnessed by the exchange of papers between Wright and Fisher, and more recently, under the mislabelled hypotheses of socalled Darwinian and Non-Darwinian evolution. Insofar the evidence from differentiation patterns in natural populations is concerned, it is helpful to develop a systematic ordering of the kinds of tests that are often used in analysing data on gene frequencies: Table 7 is an attempt to summarize the ideas involved, given data on the gene or genotypic frequencies at a number of loci ($j = 1, 2, \ldots, n$) with k_j alleles each ($i = 1, 2, \ldots, k$) for a number of populations sampled on various geographical scales (continents, regions, subregions, localities, etc.). Such data naturally lend themselves to a hierarchal analysis of variation patterns in terms of the variances, genetic distances, etc. From Table 7 we can look at a rather large number of alternative

Table 7. *Analysis of Inter-Populational Data on Allelic Frequencies (q$_{ij}$)*

A. No differentiation (s$_q^2$ = 0, heterogeneity χ^2 small, etc.) for any j, on area surveyed.

(A1) q$_{ij}$ 0 or 1; monomorphism; (a) same, or (b) different founder gene pools originally; selection in case of (b), with or without high migration; no evidence on drift.

(A2) q$_{ij}$ showing polymorphisms and homogeneous allelic frequency distribution; (a) same original stock with insufficient time having elapsed since subdivision, or no selection, (b) moderate to high gene flow, balancing off random or selective differentiation, or (c) "uniform" selection, often heterotic at a few or many loci. (a) Inconclusive, (b) assumes migration counteracting selection and drift both, but (c) favors selection for some "general pupose" genotype.
(Can (a), (b), (c) be established in any one example? Data on allelic identity, time, m, N and manipulative ecogenetic experiments are needed.)

B. Significant heterogeneity (s$_q^2$ ≠ 0, χ^2 significant) but showing no large scale geographical, regional, spatial patterns. (Depends on sampling unit, method, loci, etc.)

(B1) Original pattern lost due to recent habitat destruction leaving only remnant populations.

(B2) Niche diversity fine-scaled (local patchiness) with (a) stability or (b) instability over time.

For (a) $\overline{\Delta q}$ and s$_{\Delta}^2$ ≈ 0, but for (b) $\overline{\Delta q}$?, and observed s$_{\Delta q}^2$ greater than expected. Selection with or without drift. Multigenic associations indicate but do not prove ecotypic differentiation.

(B3) Changes for different alleles and loci showing Δq = 0, $\sigma_{\Delta q}$ as expected (inversely proportional to N$_e$, rates of fixation uniform) and covariances ≈ 0. Evidence for drift, and weak or no selection.

C. Significant heterogeneity and showing geographical, regional, clinal patterns. (Scale of pattern vs. migration scale and N$_e$.)

(C1) Selection with some environmental gradient, proportionally stronger than migration and drift multigenic associations represent ecotypic variation. Direct evidence for selection.

(C2) Central vs. peripheral or mainland vs. island patterns, associated with selection and relatively greater role of population numbers.

(C3) Mixed patterns of varying scales at different loci, but (a) showing relative stability over time, time, (b) nothing known about stability, on (c) changing over time with reasons not well understood.

Some of the wellknown examples in population genetic literature appear to fit in to more than one of these categories as follows: *Drosophila pseudoobscura* inversions and isoenzymatic variants (A2bc, C1, C2, C3ab); *Cepaea nemoralis* shell color and banding (B2, C3ab); *Mus musculus* isoenzymatic variants (A2, C1, C2, C3ab), human blood groups (B3, C1, C3ab); *Maniola jurtina* wing spotting (B2, C3); *Linanthus parryae* flower color (B2, B3); *Avena fatua* morphological markers (B2, B3, C2, C3ab); *Avena barbata* (A1, B1, C1, C2, C3ab); *Agrostis tenuis* metal tolerance (C1, C3), and so on. Both presence and absence of spatial heterogeneity require further information on population size, history, gene flow, etc. in order to test various alternative hypotheses (see Ford, 1971; Grant, 1963; Gaines and Krebs, 1971; Kimura and Ohta, 1971; Levins, 1968; Lewontin, 1967; Selander, 1970; Wright, 1969.)

outcomes, the parameters required for the proper decision making, and yet, there seem to be numerous uncertainties. Clearly, with the usual kinds of data reported in literature, in most instances we still have a large number of alternative explanations. Unfortunately, we often conclude by giving sufficient (not necessary) explanations. This leads to weak inference in evolutionary biology and until we are prepared to undertake more thorough and longterm studies it is likely to remain so. As Maynard Smith (1970) put it, we have a "choice between the Garden of Eden and Darwinism". Dobzhansky (1970) concludes that "genetic differentiation of partially isolated colonies by random drift is probably a widespread phenomenon" ... "the adaptive role of a character must be demonstrated; it cannot be assumed *a priori*". These quotes underscore the fact that most conclusions are of necessity probabilistic statements and that no-selection models should be used as null hypotheses in the analyses of gene frequency changes.

Literature

1. Dobzhansky, Th.: Genetics of the Evolutionary Process. New York: Columbia Univ. Press 1970. — 2. Ford, E. B.: Ecological Genetics, 3rd ed. London: Methuen 1971. — 3. Gaines, M. S., Krebs, C. J.: Genetic changes in fluctuating vole populations. Evolution **25**, 702—723 (1971). — 4. Grant, V.: The Origin of Adaptations. New York: Columbia Univ. Press 1963. — 5. Imam, A. G., Allard, R. W.: Population studies in predominantly selfpollinated species. VI. Genetic variability between and within natural populations of wild oats from differing habitats in California. Genetics **51**, 49—62 (1965). — 6. Jain, S. K., Marshall, D. R.: Genetic changes in a barley population analyzed in terms of life cycle components of selection. Genetica **38**, 355—374 (1968). — 7. Kimura, M., Crow, J. F.: The measurement of effective population number. Evolution **17**, 279—288 (1963). — 8. Kimura, M., Ohta, T.: Theoretical Aspects of Population Genetics. New York: Princeton Univ. Press 1971. — 9. Levins, R.: Evolution in Changing Environments. New York: Princeton Univ. Press 1968. — 10. Lewontin, R. C.: Population Genetics. Ann. Rev. Genet. **1**, 37—70 (1967). — 11. Marshall, D. R., Jain, S. K.: Genetic polymorphism in natural populations of *Avena fatua* and *A. barbata*. Nature, London, **221**, 276—278 (1968). — 12. Maynard Smith, J.: The causes of polymorphism. Symp. Zool. Soc. London **26**, 371—383 (1970). — 13. Selander, R. K.: Behavior and genetic variation in natural populations. Am. Zoologist **10**, 53—66 (1970). — 14. Wright, S.: The genetical structure of populations. Ann. Eugen. **15**, 323—354 (1951). — 15. Wright, S.: Evolution and the Genetics of Populations, Vol. 2: The Theory of Gene Frequencies. Chicago: Univ. of Chicago Press 1969.

Dr. S. K. Jain
Department of Agronomy and Range Science
University of California
Davis, California 95616 (USA)
K. N. Rai
Botanisches Institut und Botanischer Garten
der Universität Wien
Wien (Austria)