Theor. Appl. Genet. 62, 219–223 (1982)

Genetic Distance Between Populations

Part 3: Wahlund's Principle as Related to Genetic Distance and an Application

H. H. Hattemer

Lehrstuhl für Forstgenetik und Forstpflanzenzüchtung der Universität Göttingen, Göttingen-Weende
(Federal Republic of Germany)

Summary. An excess proportion of homozygous carriers of a gene arises on bulking of reproductively isolated subpopulations. This surplus of homozygotes in the mixture, measured relative to the panmictic proportion, is caused by variation of its frequencies in the respective subpopulations. It is shown that the result known as Wahlund's principle can be expressed as a simple function of genic distances among the subpopulations. If used with caution it can be employed to recognize mixtures of seed lots. The effect of bulking can be readily discriminated from that of inbreeding. It may also be distinguished from the effect of assortative mating by analysing such distances at several gene loci. Various effects of selection may disturb inference on whether a given lot is a mixture. Hence application should be confined to the dormant seed. Isozyme loci are most suitable for such checks since complete dominance occurs only rarely. Some advantages peculiar to conifer seed are discussed.

Key words: Genetic distance – Wahlund's principle – Inbreeding – Assortative mating – Selection – Seed certification

Introduction

A population is assumed to be subdivided into k subpopulations. The allelic structures of these subpopulations are assumed to differ at a diploid gene locus A in that the frequencies of an allele A_i are $p_{1i}, p_{2i}, \ldots, p_{ki}$. These differences are maintained by reproductive isolation. However, their reproductive system is panmictic with regard to traits controlled by this gene locus; hence their genotypic structures can each be characterized by Hardy-Weinberg proportions. On uniting the subpopulations, the allele A_i would attain the weighted mean frequency

$$\sum_{l=1}^{k} \mu_l p_{li} =: p_i \tag{1}$$

where μ_l is the relative size of the l-th subpopulation. Similarly the proportion of genotypes homozygous for this allele under complete isolation would become

$$\sum_{l=1}^{k} \mu_l p_{li}^2 =: P_{ii} . \tag{2}$$

The weighted variance of the frequencies of the allele A_i by definition is

$$V_i = \sum_{l=1}^{k} \mu_l p_{li}^2 - p_i^2 . \tag{3}$$

By rearranging (3) we may express the surplus of homozygous carriers arising after uniting the subpopulations but before reproduction by just this variance,

$$P_{ii} = p_i^2 + V_i , \tag{4}$$

where p_i^2 is the hypothetical proportion of homozygous carriers after one generation of panmixia in the united population. It is self-evident that this surplus is caused by reproductive isolation of subpopulations with different frequencies of the allele. This result which is due to Wahlund (1928) is demonstrated in Fig. 1.

Relation to Genetic Distance

Subdivision into two Subpopulations

Gregorius (1974) devised a measure of genic distance at a gene locus with q alleles between any two subpopulations,

$$d_{1,2} := \tfrac{1}{2} \sum_{i=1}^{q} |p_{1,i} - p_{2,i}| . \tag{5a}$$

0040-5752/82/0062/0219/$ 1.00

It measures distance between genic structures and possesses such attractive properties as

a) $d_{1,2} = d_{2,1}$

b) $0 \leq d_{1,2} \leq 1$

c) $d_{1,2} = 0$ if and only if subpopulations 1 and 2 possess identical genic structures

d) $d_{1,2} = 1$ if and only if the two subpopulations have no allele in common

e) it satisfies the triangular inequality

f) extension to genic structures in several loci is straightforward. As Gregorius (in press) shows, there are three additional such properties: d reaches its maximum if two populations have no allele in common; d is invariant to changes in frequencies of alleles not common to both populations, d is completely linear over the whole range of gene frequencies.

Equation (5 a) reduces to

$$d_{1,2} = |p_{1,i} - p_{2,i}| \tag{5b}$$

if all alleles other than A_i are lumped so that a partition of the set of alleles into A_i and its complementary set is considered. Since the variance (3) in this case of only two subpopulations amounts to

$$V_i = \mu(1-\mu)\,(p_{1,i} - p_{2,i})^2 \tag{6}$$

we may express Wahlund's effect on homozygote surplus by

$$V_i = \mu(1-\mu)\,d_{1,2}^2\,. \tag{7}$$

In this context it is irrelevant whether μ measures the proportion of the first or second subpopulation.

The variance in (7) reaches its maximum if the subpopulations are of equal size (Fig. 1).

If there exists information on the frequency of allele A_i in the two subpopulations one can infer μ from the observation of its homozygotes carriers in the aggregate population using a reformulation of (4), i.e.

$$P_{ii} = p_i^2 + \mu(1-\mu)\,d_{1,2}^2\,. \tag{4a}$$

Solving (4a) for μ yields a quadratic equation in μ with roots

$$\mu = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{4(P_{ii} - p_i^2)}{d_{1,2}^2}}\,. \tag{4b}$$

Since, also from (4a),

$$\frac{P_{ii} - p_i^2}{d_{1,2}^2} = \mu(1-\mu) \leq \frac{1}{4}$$

both roots will be non-negative and will always amount to μ and $(1-\mu)$.

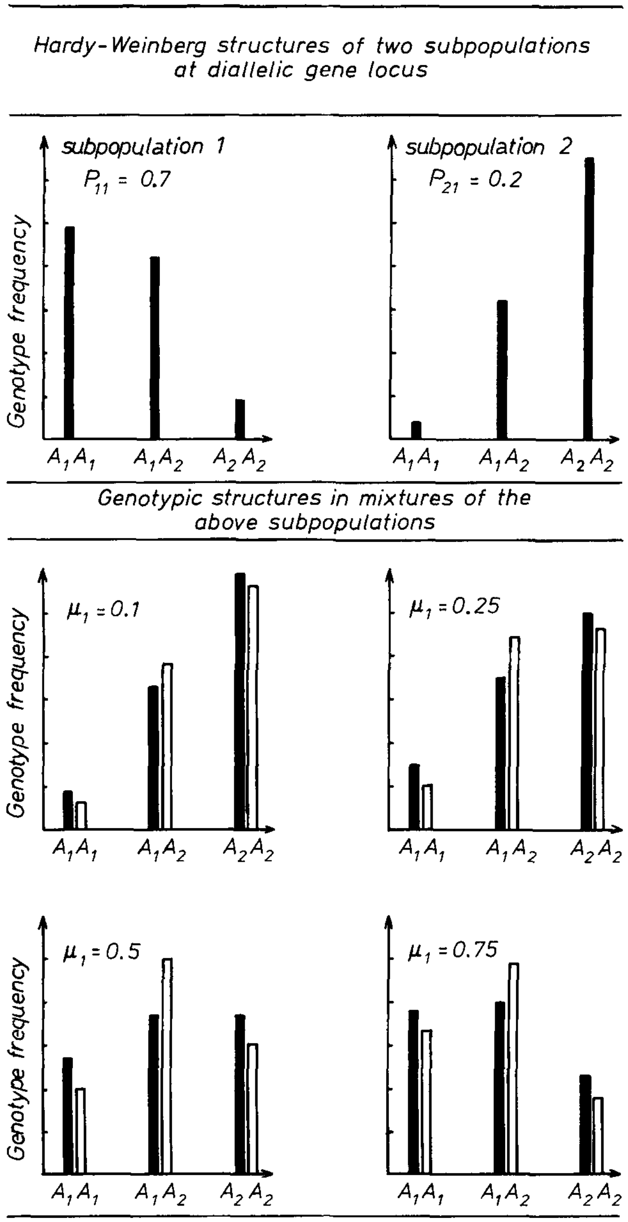

Fig. 1. Excess of homozygotes after bulking two subpopulations in Hardy-Weinberg proportions but with different allelic structures. In mixtures, solid bars denote the actual genotype structure and empty bars the corresponding Hardy-Weinberg structure; this serves as a reference by which to measure homozygote excess

Subdivision into k Subpopulations

If k subpopulations are aggregated, the squared expectation in the variance (3) can be transformed,

$$p_i^2 = \left(\sum_{l=1}^{k} \mu_l\, p_{li}\right)^2 = \sum_{l=1}^{k} \mu_l^2\, p_{li}^2 + \sum_{l=1}^{k} \sum_{\substack{m=1\\ m\neq l}}^{k} \mu_l\, \mu_m\, p_{li}\, p_{mi}$$

so that (3) becomes

$$\begin{aligned} V_i &= \sum_{l=1}^{k} \mu_l\, p_{li}^2 - p_i^2 \\ &= \sum_{l=1}^{k} \mu_l(1-\mu_l)\, p_{li}^2 - \sum_{l=1}^{k} \sum_{\substack{m=1\\ m\neq l}}^{k} \mu_l\, \mu_m\, p_{li}\, p_{mi}\,. \end{aligned} \tag{8}$$

We may now transform the first term on the right-hand side of (8),

$$\begin{aligned}\sum_{l} \mu_l(1-\mu_l)\, p_{li}^2 &= \sum_{l} \mu_l\, p_{li}^2 \sum_{\substack{m\\ m\neq l}} \mu_m = \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m\, p_{li}^2\\ &= \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m\, p_{mi}^2 = \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m\, p_{mi}^2\end{aligned}$$

so that we may partition this term. This yields

$$\begin{aligned}V_i &= \tfrac{1}{2} \sum_{l} \mu_l \sum_{\substack{m\\ m\neq l}} \mu_m\, p_{li}^2 + \tfrac{1}{2} \sum_{l} \mu_l \sum_{\substack{m\\ m\neq l}} \mu_m\, p_{mi}^2\\ &\quad - \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m\, p_{li}\, p_{mi}\\ &= \tfrac{1}{2} \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m (p_{li}^2 + p_{mi}^2 - 2\, p_{li}\, p_{mi})\\ &= \tfrac{1}{2} \sum_{l} \sum_{\substack{m\\ m\neq l}} \mu_l\, \mu_m (p_{li} - p_{mi})^2 . \qquad (9)\end{aligned}$$

We may thus equate the variance of the frequencies of an allele to half of the weighted sum of products of all squared genic distances among pairs of subpopulations, or

$$\begin{aligned}V_i &= \tfrac{1}{2} \sum_{l=1}^{k} \sum_{\substack{m=1\\ m\neq l}}^{k} \mu_l\, \mu_m\, d_{l,m;i}^2\\ &= \sum_{l=1}^{k} \sum_{m=l+1}^{k} \mu_l\, \mu_m\, d_{l,m;i}^2 . \qquad (10)\end{aligned}$$

The quantity $d_{l,m;i}$ measures only the distance between subpopulations l and m which is brought about by allele A_i. It has to be kept in mind that this relation holds only if all k subpopulations possess Hardy-Weinberg structure at the gene locus in question.

Effects on Genotypic Structure Relative to Hardy-Weinberg Proportions

After hypothetical union of the isolated subpopulations and before reproduction in the united population, the proportion of homozygous carriers of allele A_i is expected to be as in (4), while the proportion of heterozygous carriers is expected to be

$$\sum_{\substack{j\\ j\neq i}} P_{ij} = 2 p_i (1-p_i) - 2 V_i \qquad (11)$$

since the overall frequency of carriers is unaffected by subdivision.

Rearranging (11) we have

$$V_i = p_i(1-p_i) - \tfrac{1}{2} \sum_{\substack{j\\ j\neq i}} P_{ij} = p_i(1-p_i) - (p_i - P_{ii}) . \qquad (11\,a)$$

This amount is peculiar to the allele A_i and the pattern of its individual genic distances in the sense of Gregorius (1974). The influence of subdivision on the frequency of genotypes $A_i A_i$ and $A_i A_h$ ($h \neq i$) in the presence of multiple alleles in the aggregate population was investigated by Nei (1965) and Li (1969). These authors expressed the effect of random subdivision in terms of variance and covariance of an individual allele while for the detection of subdivision the present procedure given in Eqs. (4) and (11) may be more useful. Also, the measurement of the effect of isolation should be made independent of inbreeding as Li (1969) suggests. The interpretation of V_i by subdivision is not only peculiar to an individual allele, it is also based on the assumption of panmixia in the isolates as was shown in the present paper.

Discrimination of Wahlund's Effect From Effects of the Mating System

Certain properties of mating systems which are at variance with random mating induce an homozygote excess after one generation. Their effects are identical if a single gene locus is considered.

Inbreeding

Nei (1965) has pointed out that if inbreeding rather than Wahlund's effect is responsible for a surplus of homozygote carriers this can be inferred from homogeneity of estimated inbreeding coefficients among the various alleles of any gene locus, i.e. $F_{ii} = F_{ij} = \bar{F} > 0$ for any i and j (Wright 1969) so that genotype frequencies are a function of the mean inbreeding coefficient estimated in the aggregate population.

As always, the excess proportion of homozygous carriers of an allele amounts to half of the shortage of heterozygotes. Unfortunately, Nei (1965) does not explain why this simple rule in the case of multiple alleles does not hold "in the strict sense". As Gregorius (pers. comm.) points out it does hold for any gene locus no matter how this deviation from Hardy-Weinberg structure arose: The excess of homozygous carriers of an allele A_i relative to Hardy-Weinberg structure amounts to $P_{ii} - p_i^2$. Under these same conditions the corresponding shortage of heterozygous carriers of the allele is measured by

$$2 p_i (1-p_i) - \sum_{\substack{j\\ j\neq i}} P_{ij} = 2 (p_i - p_i^2 - \tfrac{1}{2} \sum_{\substack{j\\ j\neq i}} P_{ij}) = 2\, (P_{ii} - p_i^2)$$

so that the ratio of these two quantities amounts to $\frac{1}{2}$ under any genotypic structure. There is now way to make a decision on the presence of inbreeding if the gene locus has only two alleles.

If the effects of both inbreeding and subdivision are traceable the latter should reinforce the former. By reversing a statement of Brown (1979) one may say that in inbred populations there is less scope for a homozygote surplus to be caused by Wahlund's effect.

Simultaneous investigation of several gene loci should demonstrate differential genic distances but identical coefficients of inbreeding.

Positive Assortative Mating

If carriers of a genotype containing the allele A_i give a mating preference $0 < \alpha < 1$ to carriers of the same genotype, an excess of homozygous carriers of this allele is present in the immediate progeny (denoted by primes):

$$\left.\begin{aligned} P'_{ii} &= p_i^2 + \alpha[\tfrac{1}{2}(p_i + P_{ii}) - p_i^2]\,, \\ \sum_{\substack{j \\ j\neq i}} P'_{ij} &= 2p_i(1-p_i) - \alpha[2p_i(1-p_i) - \tfrac{1}{2}\sum_{\substack{j \\ j\neq i}} P_{ij}] \end{aligned}\right\} \quad (12)$$

(12) is valid for the gene loci which control the assortativity; for this group of gene loci, (12) is also equivalent to partial selfing, in which case α corresponds to the rate of self-fertilization.

The biological condition of positive assortativity in plant populations may consist of the involvement of gene locus A in the genetic control of flowering time. This is particularly true with windpollinated populations because a flower search behavior of animal pollinators might alone be sufficient to further complicate this aspect of mating. Since the weather conditions e.g. will greatly influence the timing of gamete production and thus the pattern of preferences an equilibrium will hardly ever be reached. If such an equilibrium is nevertheless assumed to exist, such as

$$\left.\begin{aligned} \hat{P}_{ii} &= p_i^2 + \frac{\alpha}{2-\alpha}\, p_i(1-p_i) \\ \sum_{\substack{j \\ j\neq i}} \hat{P}_{ij} &= 2p_i(1-p_i)\,\frac{2(1-\alpha)}{2-\alpha} \end{aligned}\right\} \quad (13)$$

the excesses of homozygote carriers of alleles are always functions of a common preference parameter α even if parameters $\alpha_{ij} \neq \alpha$ are peculiar to individual genotypes. But in contrast to partial selfing this preference rests on a phenotypic trait. Hence the excess of homozygotes at linked loci decays at rates depending on the recombination rates (Gregorius 1980), and the coefficients, F_{ij}, estimated to measure the supposed inbreeding might vary among practically all gene loci.

Though assortative mating may be classified as a mode of selection without viability or fertility parameters being different among genotypes (Gregorius, pers. comm.), its effect on genotypic structures differs markedly from the latter two selective forces.

Gene loci responsible for assortative mating may be recognized by their excess of homozygotes which may be explained by one hypothetical assortativity parameter not shared by other gene loci displaying the same phenomenon to a greater or lesser degree. Such loci may also be identified by empirical observations in samples drawn from populations that are not possibly subdivided.

Application to Seed Certification

A major application of Wahlund's principle is to indicate errors in the information provided on the label accompanying plant propagules. Excess proportions of homozygotes may imply admixtures and hence suggest comparison with the basic material stated on the label. Conventionally the identity of a commercial lot of reproductive material is established solely by tracing the route it took from collection to marketing. The purchaser's interests may thus be more efficiently accounted for by analyzing the genotypic structure of the seed or planting stock since it is the genetic information rather than information on the origin he wants to buy. The present mode of control of descent is per se non-genetic since the mating system realized in producing a given lot is largely unknown and the information even on the genealogy is incomplete. Besides, keeping record on any changes of the owner of propagules implies considerable administrative effort.

Such genetic analyses can readily be realized by the employment of diploid isozyme loci identified in the seed of many angiospermous crop plants. Since heterozygotes at such gene loci can be discriminated from homozygotes, all carriers of an allele may be readily identified. The analyses to identify a lot as a mixture may also yield information on other sources of homozygote excess important in many predominantly cross-pollinating species. Many such gene loci are active in both the seed and the fully developed plant and can thus be used to identify the seed as possible progeny of the reproductive material stated on the label. This type of control is possible also in other than diploid gene loci and in predominantly self-pollinating species; in the latter group wide-spread fixation of alleles in commercial varieties may greatly help decisions on identity.

Inspection of several gene loci (Sinnock 1975) eases such testing as does the information on the ordered genotype present in conifer seed (Müller 1976). The utilization of drastic differences of allele frequencies between geographic sources was discussed by Bergmann (1974, 1975, 1978).

Viability selection brings about genotypic structures that no longer display certain properties of the mating system. Under the incidence of high mortality as in some plant species genetic viability selection may disturb the original structure encountered in the seed (Kim 1980 for an example of a long-lived plant species). The present system of interpreting a genotypic structure (Hattemer 1982) is therefore restricted to the dormant seed, i.e. a stage of the life cycle before viability selection operated.

However, the effects of selective components of the reproductive system on genotypic structure may be unresolvably confounded with that of subdivision and subsequent bulking. In particular, as Robertson (1965) showed, sexually asymmetric differences in gamete output of genotypes may produce an excess of heterozygotes and thus counteract the process that is otherwise indicative of bulking. For instance in monoecious plant populations it is highly conceivable that opposite selection pressures operate in the two sexes because it improves the economy of energy allocation. In conifers, this condition has become testable by employing the results of Müller (1976) and it has been under intensive study since then. Other'situations of this type were described by Brown (1979). It has to be kept in mind that Wahlund's principle is defined relative to Hardy-Weinberg structure, and that deviations from panmixia disturb inference on bulking, and that a major source of inference is the estimation of (hypothetical) coefficients of inbreeding.

Future experimentation will certainly show which gene loci can be successfully employed in detecting mixtures of seed. It may be the ones that are highly polymorphic without being involved in controlling flowering time, display much variation but are neither the target of heavy fertility selection nor early zygote selection.

Acknowledgement

The author is heavily indebted to M. Ziehe and H.-R. Gregorius, Göttingen, for continuous encouragement and help. J. Roberds, Raleigh, improved the text by indicating several shortcomings, which is gratefully acknowledged.

Literature

Bergmann, F. (1974): Genetischer Abstand zwischen Populationen. II. Die Bestimmung des genetischen Abstands zwischen europäischen Fichtenpopulationen (*Picea abies*) auf der Basis von Isoenzym-Genhäufigkeiten. Silvae Genet. **23**, 28–32

Bergmann, F. (1975): Herkunfts-Identifizierung von Forstsaatgut auf der Basis von Isoenzym-Genhäufigkeiten. Allg. Forst- u. Jagdztg. **146**, 191–195

Bergmann, F. (1978): The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*) along similar climatic gradients. Theor. Appl. Genet. **52**, 67–64

Brown, A.H.D. (1979): Enzyme polymorphism in plant populations. Theoret. Popul. Biol. **15**, 1–42

Gregorius, H.-R. (1974): Genetischer Abstand zwischen Populationen. I. Konzeption der genetischen Abstandsmessung. Silvae Genet. **23**, 22–27

Gregorius, H.-R. (1980): The effect of positive assortative mating at one locus on a second linked locus. I. The genotypic structure of the offspring generation. Theor. Appl. Genet. **56**, 187–192

Gregorius, H.-R. (1980): The effect of positive assortative mating on a second linked locus. II. Limiting characteristics of genetic and genotypic structures. Theor. Appl. Genet. **57**, 17–24

Gregorius, H.-R.: A unique genetic distance. Biometr. J. (in press)

Hattemer, H.H. (1982) Genetische Untersuchungen an Forstsaatgut. Allg. Forstztg. **93**, (in press)

Kim, Z.S. (1980): Veränderung der genetischen Struktur von Buchenpopulationen durch Viabilitätsselektion im Keimlingsalter. Göttingen Res. Notes in Forest Genetics **3**, 1–87

Li, C.C. (1969): Population subdivision with respect to multiple alleles. Ann. Hum. Genet. **33**, 23–29

Müller, G. (1976): A simple method of estimating rates of self-fertilization by analysing isozymes in tree seeds. Silvae Genet. **25**, 15–17

Nei, M. (1965): Variation and covariation of gene frequencies in subdivided populations. Evolution **19**, 256–258

Robertson, A. (1965): The interpretation of genotypic ratios in domestic animal populations. Anim. Prod. **7**, 319–324

Sinnock, P. (1975): The Wahlund effect for the two-locus model. Am. Nat. **109**, 565–570

Wahlund, S. (1928): Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **11**, 65–106

Wright, S. (1969): Evolution and the Genetics of Populations, Vol. 2. The Theory of Gene Frequencies. Chicago: University of Chicago Press

Received February 17, 1982
Communicated by Å. Gustafsson

Dr. H. H. Hattemer
Lehrstuhl für Forstgenetik und
Forstpflanzenzüchtung der
Universität Göttingen
Büsgenweg 2
D-3400 Göttingen-Weende
(Federal Republic of Germany)