Published in *B. Degen, M.D. Loveless, A. Kremer 2002. Modelling and Experimental Research on Genetic Processes in Tropical and Temperate Forests. Belém, PA: Embrapa Amazônia Oriental, pp. 42-68 (reprinted with TEX)*

An integrative approach to modeling mating systems of tree populations

H.-R. Gregorius

Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen,
Büsgenweg 2, 37077 Göttingen, Germany

Keywords: matings systems, trees, modeling, estimation, functions of mating systems, adaptation, effective size, selection load, gene flow, parentage analysis, mating preferences, metapopulation structure

Abstract: The evaluation of existing approaches and the development of alternative approaches to modeling and estimation of mating system characteristics is put on a firm basis by considering the three fundamental functions of mating systems: (1) generative reproduction, (2) selection for participation in generative reproduction, and (3) combination of genetic information into genotypes. Intactness of the corresponding mating system mechanisms directly affects (1) adaptedness to current environmental conditions, (2) preservation of adaptability to future changes, and (3) preservation of adaptedness. Indicators for the intactness of these mechanisms are estimates of (1) reproductive success, (2) reproductively effective population size, and (3) selection load. The latter two are elaborated conceptually. All three indicators are integrated into the modeling and estimation of mating system characteristics by utilizing parentage distributions. A parentage distribution consists of a pool of zygotes and a group of potential parents such that each zygote has at least one parent from the group. This approach is useful for the design of models, methods of estimation and for the exploration of mating system characteristics. It is applied to measurement of the effects of mating systems on the demarcation of populations, on the reproductively effective number of maternal and paternal parents, on the amount of gene flow including its two sexual components, and on subpopulation differentiation by spatial variation in mating relations. Application of measures of mating preference as defined by the parental pairs of zygotes is shown to provide conceptually more satisfactory information on the reproductive isolation and coherence patterns that determine metapopulation structure and initiate or prevent speciation. Self-fertilization, for example, can be viewed in this context as establishing an extreme type of genealogical metapopulation structure, which is detectable with the help of estimates of "self-preference" but not of proportions of self-fertilization.

Introduction

In the present paper, an attempt will be made to bring to attention the fundamental functions of mating systems as the ultimate goals to which modeling and estimation of mating system characteristics are directed. By this, it is hoped that the comparative evaluation of existing approaches can be put on a firmer basis and that the development of alternative approaches, if desirable, is aided.

Objectives of modeling

For a start, recall that any analysis of the consequences of mating system characteristics for population development and survival must be preceded by a description and estimation of these characteristics. Modeling is required here, either as a means of predicting this development on the basis of observed characteristics, or as a means of obtaining these characteristics if no methods of direct observation are available. In the latter case, where target characteristics are not observable, a model is designed in which the target variables depend on observable variables. Calibration of this model with respect to the observations (usually by maximization of likelihood) then yields "indirect" (or model-dependent) estimates of the desired target characteristics. These estimates are only acceptable if a test of the calibrated model does not recommend its rejection. Estimation of rates of self-fertilization in trees is a well-known example of indirect estimation.

A third common area of application of models concerns the detection of equivalencies in performance between different systems. This is usually realized in terms of "effective sizes", which result from the comparison of a complex system with an "ideal" model system. In particular, this comprises all non-testable models, which yield "estimates" of model parameters after calibration with respect to observations. Such "estimates" do not refer to characteristics of the observed system and should thus be addressed as effective parameter values. An example is to be seen in F_{ST} used to "estimate" amounts of gene flow (in the form of $N{\cdot}m$) on the basis of Wright's idealized model of drift and migration. Experimental tests of this model are very difficult and possibly infeasible. $N{\cdot}m$ may, in this case, be addressed as an effective amount of migration, but it does not estimate any realized number of migrants (for a detailed criticism of the F_{ST} method as yielding indirect estimates of gene flow, see Whitlock & McCauley 1999). This category of model is thus of limited value in the analysis of characteristics of real systems.

In summary, one can distinguish three major objectives of modeling,

- ▷ to provide *testable hypotheses on causal mechanisms* and serve the *prediction* of developments and *planning* of actions,
- ▷ to enable *indirect estimation* of system characteristics, after passing a test ("model-dependent estimation"), and
- ▷ to enable the detection of *equivalencies in performance* between different systems ("effective sizes")

Mating system functions

While the above objectives are pursued to various degrees in each analytical study of mating systems, the design of the applied models frequently does not permit clear recognition of their relations to general biological functions of mating systems. Since the inherent principles can be expected to guide any modeling effort it is useful to recall briefly the three *fundamental functions of mating systems*, which are

(1) generative *reproduction*,
(2) *selection* for participation in generative reproduction,
(3) *combination* of genetic information (genes) into genotypes.

These functions of mating systems determine those adaptational capacities of populations which can be realized during the transition from one generation to the next. Consideration of the functions in experimental analyses is thus required to assess the significance of observable mating system characteristics for population development and survival. The same requirement applies, of course, to the design and parameterization of mating system models. These models always contain the observable mating system characteristics as variables. They must, however, frequently employ non-observable variables (free parameters) in order to enable the desired analysis. As was mentioned above, non-observable model variables serve in the calibration of the model and by this enable indirect estimation. They can also be varied with the aim to predict effects of certain scenarios on the three fundamental functions.

Major determinants of plant mating systems

In order to simplify reference of the following reflections to their biological basis, a short list of categories of factors is compiled, which affect plant mating systems. Among the categories most frequently considered in experimental studies, the following can be distinguished:

- (*i*) *Spatial* relations: Spatial distribution of individuals, in relation to their pollen dispersal characteristics.
- (*ii*) *Temporal* relations: Temporally varying spatial distribution patterns and activities or behavior, age of female and male sexual maturity, timing of female and male flowering, time-dependent expression of reproductively relevant phenotypes.
- (*iii*) *Phenotypic* relations: Prezygotic incompatibility or isolation mechanisms, including biochemical or physiological agents, and morphological barriers (concerning e.g. flowering phenology). With the exception of purely genetic control, expression of the relevant traits involves interactions of genotypes with environments.
- (*iv*) *Ecological* conditions: Availability and selectivity of pollinators, physical barriers to pollen dispersal, species composition.
- (*v*) *Genealogical* relations: With the exception of self-fertilization, they are rarely direct determinants of mating relations. Preferential mating among relatives in plants is mostly a consequence of limited seed and pollen dispersal, which reveals spatial rather than genealogical relations as direct determinants of mating.

With particular reference to forest tree mating systems, mixed mating in the form of selfing and random cross-fertilization (category (*v*)), preferential mating among neighbors (category (*i*)), and gametophytic and sporophytic incompatibility (category (*iii*)) have received most attention. More recently, ecological conditions (category (*iv*)) are attracting some interest because of increased concern about tropical tree species with their animal dominated pollination systems.

The adaptational context of mating systems

Intactness of a mating system

Adaptedness and adaptability of mating systems to environmental conditions can be realized only if the mechanisms performing the three basic functions of the mating system are intact. Corresponding to the three functions,

1. the mechanisms of generative reproduction are intact if sufficient numbers of offspring are produced, in the sense that the number of offspring compensates for the number of deaths (i.e. the number of offspring of a cohort over its total life span is at least equal to the cohort's initial size);
2. the mechanisms of selection for generative reproduction are intact if all of the adults' genes are represented in their successful gametes (gametes appearing in zygotes);
3. the mechanisms of combination are intact if the genotypic composition of the offspring guarantees sufficient chances for survival to adulthood and reproduction in the next generation.

Translated back into adaptational terms, this states that intact performance of the three functions implies that the mating system of a population

1. *is adapted* to the current environmental conditions,
2. *preserves the adaptability* to future environmental changes by preserving genetic variation,
3. *preserves the adaptedness* by reducing the mortality implied by the adaptational pressures on the next generation.

As opposed to mating system functions (1) and (2), adaptational forces do not directly act on function (3). The adaptational effect of the performance of function (3) is determined by the initial conditions that it provides for the performance of functions (1) and (2) in the next generation. In other words, function (3) should be performed such that the genotypes, which are adaptationally advantageous under the environmental conditions of the next generation, are produced at sufficient frequencies.

The mechanisms of the mating system are therefore impaired in the case of insufficient offspring production (function (1)), participation of only a small fraction of adults in the offspring production (function (2)), or in the case of excessive formation of adaptationally inferior genotypes (function (3)).

Example: inbreeding depression

An example of the performance of mating system function (3) is provided by the degree of self-fertilization, in combination with homozygote disadvantage, which is frequently addressed as inbreeding depression. For given allele frequencies, the share of adaptationally inferior genotypes increases with increasing degree of selfing. Over the generations, this share will gradually lessen for dominant gene action, until a selection-mutation equilibrium is reached. When starting with a high share of inferior genotypes, the following reduction in population size could be so drastic that the implied genetic drift effects could entail substantial losses of adaptationally important variation in the genetic background. Under such conditions, the intactness of the mating system becomes manifest in the effects of the degree of selfing on the average survival and reproduction of the offspring generation.

Example: mode of pollination

Performance of functions (1) and (3) can be affected simultaneously, for example, by the mode of pollination. According to Lloyd (1979), three such modes can be distinguished: prior, delayed, and competing self-pollination. Ziehe & Gregorius (1988) demonstrated that each of these modes affects the pollination efficiency (function (1)) and the degree of self-fertilization (combination of genes into genotypes, function (3)) differently. In particular, delayed self-pollination may increase pollination efficiency in a supplementary way if cross-pollination was insufficient or failed as a result of low population density or of colonization events. Note that the assumptions of the classical mixed mating model (random cross-fertilization, fixed ovule selfing proportions, all ovules have the same chance to be fertilized) postulate the performance of the three fundamental functions rather than explain how they are affected by certain mating relations.

Example: evolutionary effects of mating systems

The adaptational pressures on populations may reinforce the evolution of reproductive separation (the “Wallace effect” after Wallace 1889) or of reproductive coherence (Steiner & Gregorius 1997) between unlike genetic types. Separation and coherence correspond to preventing and enhancing heterotypic matings. Consequently, reproductive separation initiates speciation, and reproductive coherence stabilizes the population as a reproductive community. In both cases the losses due to selection are reduced and thus the adaptedness to the respective environmental conditions improved. Again, intactness of the mating mechanisms refers to the performance of function (3).

Example: subpopulation differentiation

The concepts of reproductive separation and reproductive coherence are also relevant at the population or metapopulation level. Adaptation to spatially heterogeneous environments requires a certain degree of reproductive isolation in order to limit the formation of adaptationally disadvantageous genotypes (function (3)). Depending on the type of environmental heterogeneity (spatial, temporal, etc.), the causes for the isolation may fall into any of the above-listed categories (*i*) to (*iv*) of determinants of mating. Reproductive isolation leading to limited reproductive neighborhoods may be a prerequisite for the evolution of local adaptations. Selfing need not contribute to the formation of reproductive neighborhoods, but it directly affects the selection loads within these neighborhoods.

Indication of the adaptational status of a mating system

It remains to demonstrate the practicability of the present approach of tracing back the adaptational status of mating systems to the intactness of their mechanisms performing the fundamental functions. This requires us to specify the indicator variables for quantification of the intactness of mechanisms of the mating system, such that they attest fulfillment of the criterion of population survival. Reflecting the three fundamental functions and their conditions of intact performance, the above exposition suggests as primary indicator variables

(1) the *reproductive success*, as defined by the number of successful gametes produced per member of a cohort,

(2) the *reproductively effective population size*, as defined by the effective number of population members contributing to the zygotes produced in a specified period of time, and

(3) the *cohort selection load*, as referred to the zygotes that established the cohort and the number of successful gametes of each cohort member (fitness). The cohort selection load is then defined by minimum reduction of the reproductive capacity of a cohort that is required to arrive at the actual differences in production of successful gametes between types (a generalized concept of selection load including survival and reproduction is introduced in Appendix I).

Even though the experimental verification of indicator (1) simply requires counts of cohort members and their offspring, these counts may be difficult to obtain in iteroparous organisms, since offspring cannot be unambiguously assigned to a single cohort. In such cases, model-dependent methods of estimating life table data must be applied, which are based on observations from different cohorts.

The example of inbreeding depression can again be used to demonstrate the effects of the mating system via self-fertilization on the cohort selection load (indicator (3)). In combination with low reproductively effective population sizes (indicator (2)), large selection loads resulting from unbalanced degrees of selfing can accelerate the loss of adaptational capacity by the loss of genetic variation.

Among the three primary indicators, the third is probably most difficult to study comprehensively in long-lived organisms like trees. It can, however, be very informative to consider defined phases of the reproductive cycle separately. As an example, this is the case for early stages, where the drastic reductions of the population size following seed production allow for strong selective adaptation. In this case, a considerable fraction of the cohort selection load is attributable to early developmental stages. Postzygotic incompatibility constitutes an important special form of this load. It is therefore useful to restrict studies of the effects of mating and viability selection on the selection load to special phases.

A more comprehensive but strongly model-dependent idea of the overall cohort selection load can be obtained from a comparison of adult trees with their seed production. Assuming that the predominant characteristics of the mating relations do not change essentially over the generations, the adult genotypic structures can be considered to have resulted from viability selection that acted on zygotic frequencies which were similar to those observable among the offspring of these adults. Thus, reverting the actual succession of mating and selection in this way, the assumption permits computation of selection loads.

The requisites for the determination of selection loads are genotypic frequencies at two successive stages, the first of which is close to the zygotic stage. While this is experimentally feasible in most situations, the requisites for the direct determination of indicator (2), the reproductively effective population size, are more difficult to realize. The reason is that for direct determination, methods are required which allow the identification of the parents of each offspring in a sample (as detailed in the next chapter on "parentage distributions"). With the perfection of methods of DNA-analysis, the chances for obtaining such direct information are likely to improve considerably. Yet, since reproduction means identical multiplication of individual genetic information, model-dependent methods of estimating reproductively effective sizes in plants will always be required.

The commonly applied methods of estimating effective population sizes rely on quite restrictive model assumptions that are hard to verify or test (see e.g. Schoen & Brown 1991 for an application of such a method). This problem is aggravated by the fact that in many studies, the characteristics for which effective sizes are defined (such as inbreeding, variance, drift, reproduction, etc.) are not clearly stated. In fact, reproductively effective sizes are occasionally subsumed under some of these characteristics, and it appears that the above definition is not yet explicitly applied in theoretical or experimental work.

Some mating system characteristics may affect several indicators of intactness simultaneously. This is true for the above-mentioned modes of pollination, which affect indicators (1) and (3) via determination of the reproductive success and the selection load. In more complex situations like this, the question of how mating systems affect the indicators of intactness can frequently be answered only with the help of simulation scenarios, based on models.

Characterizing mating systems by parentage distributions

When a decision is to be made on the mating system characteristics to be studied, the above explanations suggest that their potential effects on the intactness of mechanisms of the mating system should be taken into consideration. For the applied models, this requires that their design and parameterization should enable inferences on the status of intactness of the addressed mechanisms. Since the basis for these inferences is provided by the three primary indicator variables of intactness, the model should supply information on the differential reproductive success as part of the mating process. This can be achieved in two ways. The direct approach consists of designing a model to yield estimates of the indicators, after calibration for the observations. Otherwise, the model design should at least provide for results which can be used in other models that produce the indicator variables under realistic scenarios. Estimates of a proportion of self-fertilization could, for example, be obtained with the help of a model that contained no assumptions on the numbers of zygotes produced by the various genotypes (this is true for the classical mixed mating model, for example). Another model, in which such numbers are explicitly taken into account, could be compatible with the former model, such that the selfing estimates could be adopted and an analysis of intactness of the mating system mechanisms could be carried out. In any case, a clear concept of the kind of observations that would be useful may account for all of these aspects.

Inclusion of intactness aspects into the analysis can be realized in an ideal manner, if experimental and model designs focus on each individual offspring as representing a unique and successful mating event. The characterization of a mating episode would thus be based on offspring (ideally zygotes) as units of observation, where for each offspring two "traits" are scored, one indicating its maternal and the other of its paternal parent. These observations specify a *parentage distribution* for each collection of offspring. The distribution refers to any frequency distribution on the collection of offspring (see Table 1). The parentage distribution thus summarizes all of the information relevant for the estimation of frequencies of mating types and reproductive successes.

Table 1

	♂–parent				
♀–parent	1	2	3	4	...
1	Z_{11}	Z_{12}	Z_{13}	Z_{14}	...
2	Z_{21}	Z_{22}	Z_{23}	Z_{24}	...
3	Z_{31}	Z_{32}	Z_{33}	Z_{34}	...
4	Z_{41}	Z_{42}	Z_{43}	Z_{44}	...
...	...	...	...	...	...

Parentage distribution with Z_{ij} := number of zygotes with the i-th individual as maternal and the j-th individual as paternal parent. The Z_{ii}'s of the diagonal represent numbers of zygotes resulting from self-fertilization. If the i-th individual is a female, then $Z_{ji} = 0$ for all j, i.e. the i-th column in the table consists of zeros. Dioecious species are therefore characterized by the fact that if the i-th row contains positive elements then the i-th column consists only of zeros and vice versa.

Paternity analysis (Hamrick & Schnabel, 1985) is a well known example for the utilization of parentage distributions. This method accounts for the fact that in plants seeds are usually collected before dispersal from each of a sample of individuals. In this case the maternal parent of each offspring is known and the paternal parent is to be inferred with the help of gene markers. The relevant methods are mainly based on paternity exclusion complemented by likelihood estimation procedures of paternity (see e.g. Weir 1996, p.209ff). It is well known that the precision of the resulting inference depends heavily on the available marker, the sample of potential paternal parents and the samples of seeds. Models have a substantial part in the likelihood estimation procedures, even though they are frequently not explicitly mentioned (mostly concerning free recombination, Mendelian inheritance, stochastic independence among loci, absence of postzygotic selection, or random mating). Particularly the assumption of random mating in the model-dependent estimation of parentage is problematic, if the estimates are used in an analysis of mating relations, because of the danger of circular reasoning.

The set of zygotes on which an analysis of mating system characteristics is to be based depends on the problem to be studied. In the following sections, this will be demonstrated by addressing a few problems of elementary significance. Suggestions for an integrative approach to their treatment with the help of parentage distributions will be made.

Demarcation of populations

Even if the totality of a population's seed production could be representatively sampled prior to dispersal, only the maternal contribution to each seed can be definitely stated to have originated from this population. The paternal contribution could result via pollen immigration from other populations. Moreover, a pollen grain produced by a population member could fertilize an ovule produced by the member of another, neighboring population. Seed dispersal could bring the resulting seed back into the vicinity of its paternal parent. Strictly speaking, this raises the question as to the population to which such a seed should be assigned and thus brings to attention the demarcation of a population as a reproduction community. Since a reproduction community is, in turn, defined by the mating relations of its members, mating systems can be conceived of as fundamental determinants of populations.

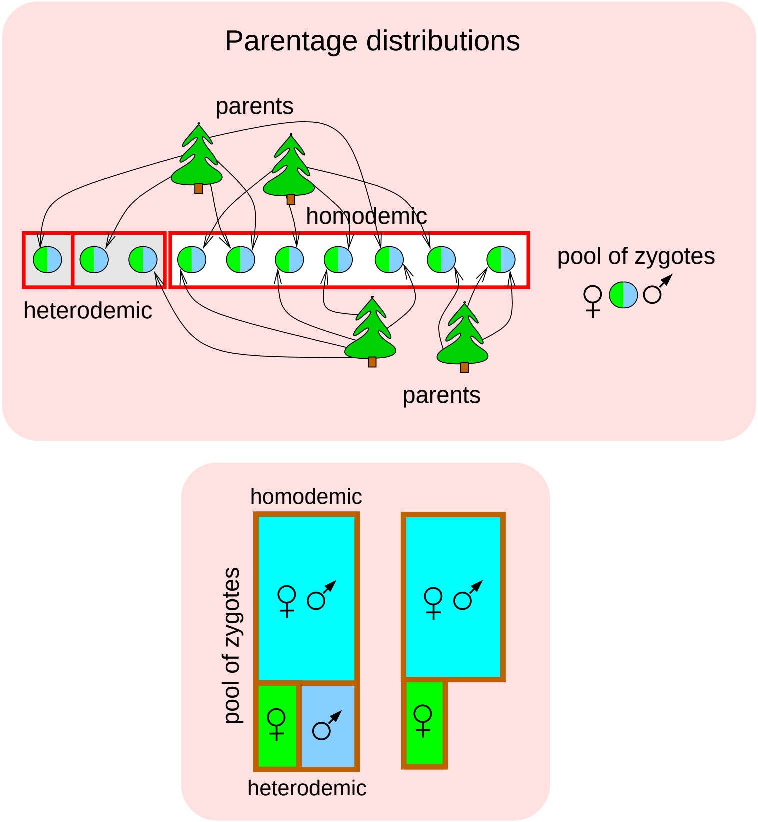
To approach this problem, the above notion of parentage distributions will be generalized to include the situation of a group of potential parents, together with all zygotes with at least one parent from this group. All of these zygotes, but no others, result from mating relations realized by the members of the group. They must therefore be part of the group's mating system, irrespective of the place where they become established (grow). To simplify wording, zygotes with both parents from the group will be called "homodemic" with respect to this group, and zygotes with only one parent from the group will be called "heterodemic" (for an illustration see top of Figure 1). There is probably wide consent to call the group a population if all zygotes are homodemic with respect to the group. Yet, since populations are also generally accepted to be open systems, sufficiently small fractions of heterodemic zygotes are counted as a result of gene flow by mating into the group, without questioning its status as a population.

With increasing fractions of heterodemic zygotes the population concept does gradually become blurred and no clear delineation can be made. It could therefore be meaningful and is probably closer to reality, if one specifies the *degree to which a group behaves as a closed population* by the fraction of homodemic zygotes among all zygotes (homo- and heterodemic) of the group. This fraction becomes 1 for a completely closed population (or reproductively isolated group) and it reaches a value of 0 if the members of the group mate only with individuals outside the group. A trivial example of a group with a zero degree of its population status is provided by any set of males in a dioecious species. To realize a positive degree, some females are to be added to this group.

Gene flow by mating

For a given population, its gene flow by mating is described by the set of heterodemic zygotes. To enable a formal representation, denote by Z the number of zygotes with at least one parent from the population, by Z_{hom} and Z_{het} the numbers of homodemic and heterodemic zygotes, and by $Z_{het}^{\female}$ and $Z_{het}^{\male}$ the numbers of heterodemic zygotes with maternal and paternal parent, respectively, from the population. Then $Z_{het} = Z_{het}^{\female} + Z_{het}^{\male}$ and $Z = Z_{hom} + Z_{het}$. The fraction of zygotes resulting from gene flow thus equals Z_{het}/Z (also note that this equals the complement of the above degree Z_{hom}/Z to which a group behaves as a closed population, see Figure 1). This obvious measure of the overall amount of gene flow by mating does not, however, underly the common measures of gene flow. The reason is to be found in the fact that in the common concept an offspring is assigned

Figure 1



Top: Illustration of homodemic and heterodemic matings in parenatge distributions. Bottom: Illustration of amounts of gene flow by mating considering both maternal and paternal parents (left) and considering only maternal parents (right) among the heterodemic matings.

to the population to which either its paternal or (mostly) its maternal parent belongs. Only from this point of view can gene flow by mating be conceived of as being directed into or out of a population. If neither gametic sex can be assigned a sessile role, gene flow cannot be analyzed for its direction.

In plants, seeds are almost exclusively assigned to the population of their maternal parents. Consequently, gene flow into a population by mating is measured by the fraction of heterodemic seeds, but excluding all those heterodemic seed with paternal parent from within the population and maternal parent from outside the population (i.e. $Z^{♂}_{het}$, see bottom of Figure 1). With the above notation, this fraction amounts to $Z^{♀}_{het}/(Z - Z^{♂}_{het})$. Clearly, this measure of gene flow can be used to describe external mating relations of population members functioning as female parents. The external mating relations of population members functioning as male parents, which would represent the gene flow out of the population, are here completely ignored. This marks the strong bias observable in the great majority of experimental and theoretical studies (see e.g. the review of Adams & Birkes, 1991, which still provides a good account of the prevailing basic approaches to modeling and estimation of mating system characteristics of forest tree populations; Willson, 1994, reviews sexual selection in plants as being female governed; Gregorius et al., 1987, demonstrate the differential effects of both sexes on the measurement of self-fertilization).

The tacit assumption that female parents are the predominant determinants of plant mating systems still awaits an experimental verification. Concerning external mating relations, this would require at least a comparison between maternal and paternal parents with respect to their amounts of heterodemic zygotes. The fraction of heterodemic zygotes characterizes the extent to which the three fundamental functions of mating systems are realized through reproductive contacts with individuals from outside the group. Therefore, and because of the differences between genetic information transmitted by male and female gametes (chiefly concerning extranuclear information such as resides in mitochondria and plastides), it is important to have reliable information about a possible asymmetry between the sexes in their external mating relations ($Z^{♀}_{het} \neq Z^{♂}_{het}$). The development of methods for the estimation of amounts $Z^{♂}_{het}$ of external matings by paternal parents is a big challenge.

In fact, the fraction Z_{het}/Z of overall gene flow by mating always exceeds the female oriented fraction $Z^{♀}_{het}/(Z - Z^{♂}_{het})$ as can be taken from

$$\frac{Z_{het}}{Z} - \frac{Z^{♀}_{het}}{Z - Z^{♂}_{het}} = \frac{Z^{♂}_{het} \cdot Z_{hom}}{Z \cdot [Z - Z^{♂}_{het}]}$$

Hence, even strong (but with the presently available experimental means, hardly verifiable) assumptions such as sexual symmetry among heterodemic mating relations ($Z^{♀}_{het} = Z^{♂}_{het}$) cannot compensate for this difference in the measurement of gene flow by mating.

Gene flow, in the sense of heterodemic matings, involves intactness considerations through its effects on the reproductive success of the population members, on the reproductively effective population size, and on the genotypic structure among the offspring. The assessment of reproductive success and effective size is complicated by the fact that heterodemic matings involve only one parent from the population. Gene flow by mating may thus affect the intactness performance of all three fundamental functions of mating systems.

Reproductively effective number of parents

There are two approaches to the reproductively effective numbers of parents. One approach focuses on a group of potential parents, the reproductive output of which is represented by the totality of their successful gametes. In terms of zygotes, the totality of successful gametes is contained in the pool of all zygotes with at least one parent from the group of potential parents. An effective number of parents is thus defined for a specified group of potential parents. The other approach focuses on a pool of zygotes as the reproductive output of their parents. An effective number of parents is here defined for a specified pool of zygotes. The following derivations will be formulated such that they apply equally to both approaches.

Again denoting by Z the number of zygotes under consideration, it follows that the effective number of parents of these zygotes cannot exceed a number of $2{\cdot}Z$. More specifically, let Z_{ij} as specified in Table 1, so that $\sum_{i,j} Z_{ij} = Z$. The number of zygotes produced by self-fertilization of the i-th parent equals Z_{ii}. If all parents would reproduce solely by self-fertilization, the effective number of parents could not be more than Z, which is half of the maximum number realizable without self-fertilization.

It follows that the i-th parent occurs $\sum_j (Z_{ij} + Z_{ji})$ times as a contributor of a gamete to the pool of zygotes under consideration. In terms of relative frequencies, the i-th parent therefore has a share of $g_i := \sum_j (Z_{ij} + Z_{ji})/(2{\cdot}Z)$ among all gametes contributed to the zygotes. Note that $\sum_i g_i = 1$ since all parents of the pool of zygotes are taken into consideration. These relative frequencies allow us to relate the concept of effective number to that of diversity. One of the most frequently applied measures of diversity is $v_2 = (\sum_i g_i^2)^{-1}$, which, considering the definition of the g_i's, *specifies the reproductively effective number of parents in terms of the diversity of parents in their contribution of gametes to the pool of zygotes.* The measure v_2 is one from a continuum of measures v_a (given by $v_a = (\sum_i g_i^a)^{\frac{1}{1-a}}$), where the index a runs from 0 to ∞, and where v_0 equals the number of types (parents) found in a collection and $v_\infty = (\max_i g_i)^{-1}$. For a given frequency distribution, v_a decreases with increasing a, so that v_0 and v_∞ constitute the largest and smallest diversity measure in this continuum (Gregorius 1978).

The smallest measure v_∞ is distinguished by a property which has particular intuitive appeal as a measure of the reproductively effective number. Considering the parent that contributes the most to the pool of successful gametes as a reference for effectiveness, it is of immediate interest to know how many of such parents would have sufficed to produce all of the gametes of the pool. This number would ideally reflect the notion of a number of reproductively effective parents. In fact, if G denotes the overall number of successful gametes, then $G \cdot \max_i g_i$ equals the maximum number of successful gametes contributed by a single parent. If each parent would contribute this number to the pool of successful gametes, it would require $G/(G \cdot \max_i g_i) = (\max_i g_i)^{-1}$ such parents to account for all the gametes.

The effective number of parents alone makes an incomplete statement as to the intact performance of the second fundamental function of mating systems. It remains to relate the pool of zygotes and their parents to a group of individuals considered as a population of potential parents, as is pointed out in the first of the two above-mentioned approaches. These potential parents must, of course, comprise at least one of the actual parents for each zygote from the considered pool of zygotes. By this, it is guaranteed that all zygotes are offspring of members of the group, with the possibility that not all zygotes have both parents from the group (existence of heterodemic zygotes) and that not all of the successful gametes of the group are represented in the pool of zygotes. The reproductively effective size of the group is defined relative to the considered pool of zygotes, and it equals the above effective number of parents only if all zygotes are homodemic for the group.

Otherwise, each heterodemic zygote is represented by only one parent from the group, so that the maximum number of parents from the group contributing to the zygotes reduces from $2{\cdot}Z$ to $2{\cdot}Z - Z_{het}$. Then the i-th group member has a share of $g_i = \sum_j (Z_{ij} + Z_{ji})/(2{\cdot}Z - Z_{het})$ among all gametes contributed to the zygotes by the group. Note that the subscript j runs over the whole set of parents of the zygotes (thus including parents from outside the group), while the subscript i refers only to group members. Hence, summation of the g_i's only over group members yields 1. The above indices v_a, when applied to these g_i's, are again possible measures of the reproductively effective number of group members.

The same principles apply to the determination of sex-specific reproductively effective numbers. Thus, the reproductively effective number of maternal parents rests on the fractions $g_i^{♀} = \sum_j Z_{ij}/(Z - Z_{het}^{♂})$ of gametes contributed by the i-th group member to all successful female gametes of the group. The pertinent fractions for the reproductively effective number of paternal parents are $g_j^{♂} = \sum_i Z_{ij}/(Z - Z_{het}^{♀})$, where j refers to the j-th group member.

The above explanations apply to any given group of individuals and pool of zygotes, provided each zygote from the pool has at least one parent from the group. This allows us to treat a large variety of situations in terms of reproductively effective numbers of parents. For example, if the seed of a single tree is to be analyzed with respect to its effective number of paternal parents, this can be done on the basis of the above frequencies $g_j^{\male}$. In this case $g_j^{\male} = Z_{ij}/Z$ and $Z_{het}^{\female} = 0$, since only one maternal parent (individual i) is considered and since all paternal parents of the zygotes are included in the group of potential paternal parents. This can be extended to any group of potential parents, in combination with the pool of all zygotes with maternal parent from the group. It is also possible to invert the point of view by consideration of the pool of all zygotes with paternal parent from the group. This case will, however, be hardly possible to study in plants, because the totality of ovules fertilized by a pollen parent can normally not be sampled.

Population structure due to mating preferences in continuous populations

The structure of metapopulations is closely related to subpopulation differentiation due to mating preferences. This type of differentiation can again most consistently be analyzed if parentage distributions can be assessed. Parentage distributions are to be specified for the differentiation criterion to be studied. For example, if the differentiation criterion is membership of locally defined groups, each offspring is characterized by the group membership of its maternal and paternal parent. Since location is a spatial characteristic, the above category (i) of mating system determinants applies. Parentage distributions are then specified by the frequencies of offspring with parents belonging to the same group (homodemic offspring) or belonging to different groups (heterodemic offspring). If external matings are considered not to contribute to population subdivision, only offspring with both parents from the total population (homodemic for the total population) enter the analysis.

Subpopulation differentiation due to preferential mating among members of the same group can then be inferred by comparison of the frequencies of homodemic and heterodemic offspring, since these correspond to matings within and between groups. If no obvious subdivision of the population into disjoint groups is observable (i.e. if the population is continuous), an analysis of mating systems is usually aimed at the detection of relationships between spatial distance and mating. A mere analysis of frequency distributions of mating distances (frequencies of zygotes with given spatial distance between their parents), however, may not be satisfactory. Spatially heterogeneous distribution of potential mating partners may feign preferential mating among neighbors simply because of clumped occurrence. The situation is similar to the bias of spatial autocorrelation analyses by clumped spatial distributions of population members. Such pitfalls can be avoided when the analysis is based on a clear concept of mating preferences, as will be demonstrated in the following section.

Spatial distance and individual mating preferences

As a rule, preferential mating among spatial neighbors cannot produce discrete subpopulation structures if neighborhood is distributed more or less evenly in space. In this case, an analysis of subpopulation differentiation may be inadequate, since no *a priori* discrete subpopulation structure can be identified. A more adequate approach to studies of population structure is then suggested by an analysis of the mating preferences of each parent as a function of the distance between the parent and its mates.

More concretely, consider all zygotes which have the i-th individual as female parent, the frequency of which is $\sum_l Z_{il}$. The j-th potential male parent appears at a proportion $Z_{ij}/\sum_l Z_{il}$ as contributor to these zygotes, and for each such male parent its distance from the reference female parent i is recorded. Note that individuals are regarded as potential parents only to the degree to which they contribute to the pool of zygotes. In order to detect special preferences with respect to distance, all paternal contributors to the total pool of zygotes have to be drawn upon, together with their distances from the reference female parent, for comparison. On this set of potential mates, in which the j-th paternal parent is represented with a proportion $\sum_k Z_{kj}/Z$, the mating preferences are based. The preference of the female parent for a particular mate is then obtained by computing (a) the frequency of this mate among all mates of the female and (b) the frequency with which the mate occurs among the paternal contributions to the total pool of zygotes. Division of (a) by (b) yields the desired measure of preference. The mating preference $U_{j\triangleleft i}$ of the i-th maternal parent for the j-th paternal parent is therefore given by

$$U_{j\triangleleft i} = \frac{Z_{ij}}{\sum_l Z_{il}} \Bigg/ \frac{\sum_k Z_{kj}}{Z} = \frac{Z_{ij}\cdot Z}{[\sum_l Z_{il}]\cdot[\sum_k Z_{kj}]}$$

(for the concept of mating preferences see Gregorius 1989). For each mate of the reference female parent, its mating preference and its spatial distance are thus known, and this permits an analysis of neighborhood mating by plotting distance against mating preference. The resulting graphs are directly interpretable (consult Figure 2 for an illustration of isotropic – i.e. independent of direction – mating preferences).

Figure 2

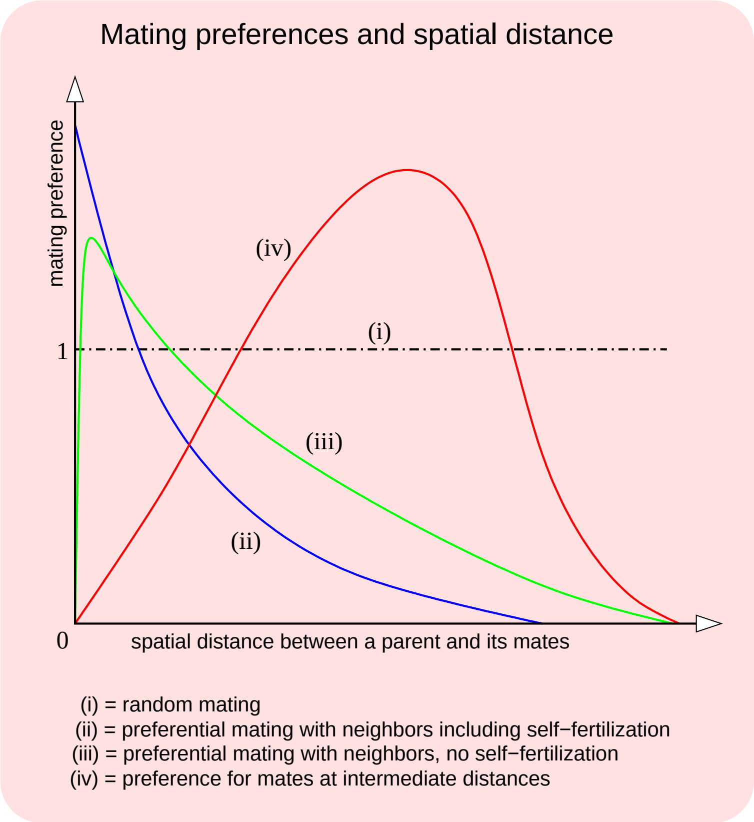


Illustration of isotropic mating preferences as a function of distance between mating partners for different mating systems.

This procedure can be repeated for each female parent and would provide an impression of the degree to which spatial distance determines female mating preferences. The same can be done with the mating preferences of male parents, in which case $U_{i \triangleleft j}$ denotes the preference of the j-th paternal parent for the i-th maternal parent. The symmetry of the preferences, i.e. $U_{i \triangleleft j} = U_{j \triangleleft i}$, follows directly from the above definition of mating preferences. This symmetry is a consequence of considering individuals as potential mating partners, to the degree to which they contribute to zygotes.

In such an analysis, spatial distance between parents can in fact be replaced by any other measure of difference between parental characteristics, without having to change the principle of the analysis. This is easily realized for the categories (*iii*) and (*v*) of mating system determinants, when considering differences between parental pairs with respect to their phenotypes, genotypes, or with respect to their common ancestry, measured in terms of coefficients of kinship.

A problem frequently arising in parentage analyses of co-sexual plants consists in the lack of means to distinguish ovule from pollen contributions to zygotes or embryos. Parents of zygotes may still be identifiable, but they cannot be distinguished with respect to their maternal and paternal functions. In this situation observations are restricted to the symmetric frequencies $Z^{\circ}_{ij} := Z_{ij} + Z_{ji}$ for $i \neq j$ and $Z^{\circ}_{ii} := Z_{ii}$. Obviously $Z^{\circ}_{ij} = Z^{\circ}_{ji}$ and $\sum_{i \leq j} Z^{\circ}_{ij} = Z$. Under these restrictions, the frequency with which the i-th individual contributes as (maternal or paternal) parent to the zygotes equals $\sum_j (Z_{ij} + Z_{ji}) = \sum_j (1 + \delta_{ij}) Z^{\circ}_{ij}$, where $\delta_{ii} = 0$ and $\delta_{ij} = 1$ for $i \neq j$. Therefore, among all zygotes with the i-th individual as parent, a proportion $Z^{\circ}_{ij} / \sum_k (1 + \delta_{ik}) Z^{\circ}_{ik}$ has the j-th individual as second parent, and this parent appears at a proportion of $\sum_k (1 + \delta_{jk}) Z^{\circ}_{jk} / (2 \cdot Z)$ among the $2 \cdot Z$ contributions of all parents to the set of zygotes. By the above definition, one now arrives at a mating preference $U_{j \triangleleft i}$ of the i-th for the j-th parent, given by

$$
\begin{aligned}
U_{j \triangleleft i} &= \frac{Z^{\circ}_{ij}}{\sum_k (1 + \delta_{ik}) Z^{\circ}_{ik}} \bigg/ \frac{\sum_k (1 + \delta_{jk}) Z^{\circ}_{jk}}{2 \cdot Z} \\
&= \frac{2 \cdot Z^{\circ}_{ij} \cdot Z}{\left[\sum_k (1 + \delta_{ik}) Z^{\circ}_{ik}\right] \cdot \left[\sum_k (1 + \delta_{jk}) Z^{\circ}_{jk}\right]}
\end{aligned}
$$

All of the above principles for an analysis of spatial distance as a determinant of sex-specific mating preferences apply equally to the symmetrical preferences.

Discrete population structure, due to mating preferences

Treatment of a broader spectrum of problems requires extension of the measurement of mating preferences to arbitrary traits of the parents. The individual parent, which was focused on in the last chapter, is then replaced by the set of all parents with the same trait state. In principle, any of the categories (*i*) to (*v*) of mating system determinants can be treated on this basis. In most cases, the traits of interest are of a discrete type or can be classified into such types so that distinguishable groups or demes of potential parents can be specified. Matings (as determined by the parentage distribution) can then again be characterized by the affiliation of mating partners (parents) to groups, and the frequencies of matings within and between the groups (homodemic and heterodemic matings) can, for example, form the basis for an analysis of differentiation among groups with respect to their mating relations. To provide a conceptual basis for such an analysis, each zygote is now considered to express two traits, specified by properties of its maternal and paternal parent. Variables X and Y will be used to identify the maternal and paternal trait, respectively.

The basic frequencies which are required for the computation of mating preferences are then given by the frequency $P(Y{=}y \,|\, X{=}x)$ of zygotes with paternal parents (Y) of type y among all zygotes with maternal parents (X) of type x and the frequency $R(Y{=}y \,|\, X{=}x)$ of paternal parents of type y among all potential mates of maternal parents of type x. With this notation, the mating preference of maternal parents of type $X{=}x$ for paternal parents of type $Y{=}y$ reads

$$U_{Y=y \triangleleft X=x} = \frac{P(Y{=}y \,|\, X{=}x)}{R(Y{=}y \,|\, X{=}x)}.$$

The inverse quantities $P(X{=}x \,|\, Y{=}y)$, $R(X{=}x \,|\, Y{=}y)$ and $U_{X=x \triangleleft Y=y}$ are defined analogously.

If, as was done in the last chapter for individual mating preferences, the potential mates are equated to their overall maternal and paternal contributions to the pool of zygotes, then $R(Y{=}y \,|\, X{=}x) = P(Y{=}y)$ and $R(X{=}x \,|\, Y{=}y) = P(X{=}x)$, where $P(Y{=}y)$ and $P(X{=}x)$ equal the frequencies or probabilities of zygotes with paternal parent (Y) of type y and maternal parent (X) of type x, respectively.

Given the frequencies of the potential mating partners, it follows that the mating preference $U_{Y=y \triangleleft X=x}$ is bounded from above by $R(Y{=}y \,|\, X{=}x)^{-1}$ since $P(Y{=}y \,|\, X{=}x) \leq 1$. This upper bound is reached if maternal parents of type $X{=}x$ mate exclusively with males of type $Y{=}y$, which is indeed the highest preference an individual can realize among its potential mates. Hence, to make more apparent the concept of preference and rejection of potential mates and to allow its quantification over the whole range from complete preference via indifference to complete rejection, it is desirable to normalize the measures U such that they vary symmetrically between $+1$ and -1. Symmetry around 0 is required to enable comparison of the extents of rejection and preference. This is achieved by the following normalization

$$\hat{U}_{Y=y \triangleleft X=x} := \begin{cases} U_{Y=y \triangleleft X=x} - 1 & \text{if } U_{Y=y \triangleleft X=x} \leq 1 \\ \dfrac{U_{Y=y \triangleleft X=x} - 1}{R(Y{=}y \,|\, X{=}x)^{-1} - 1} & \text{if } U_{Y=y \triangleleft X=x} > 1 \end{cases}$$

As desired, $\hat{U}_{Y=y \triangleleft X=x} = -1$ for complete rejection of $Y{=}y$ mates by $X{=}x$ females ($U_{Y=y \triangleleft X=x} = 0$), $\hat{U}_{Y=y \triangleleft X=x} = 1$ for exclusive mating of $X{=}x$ females with $Y{=}y$ males, and $\hat{U}_{Y=y \triangleleft X=x} = 0$ for indifference of $X{=}x$ females towards $Y{=}y$ males (random mating, $U_{Y=y \triangleleft X=x} = 1$).

$\hat{U}$ can be written in a more compact form if one considers that for $U_{Y=y \triangleleft X=x} > 1$, one obtains

$$\begin{aligned} \hat{U}_{Y=y \triangleleft X=x} &= \frac{R(Y{=}y \,|\, X{=}x) \cdot (U_{Y=y \triangleleft X=x} - 1)}{1 - R(Y{=}y \,|\, X{=}x)} \\ &= \frac{P(Y{=}y \,|\, X{=}x) - R(Y{=}y \,|\, X{=}x)}{R(Y{\neq}y \,|\, X{=}x)} \\ &= \frac{R(Y{\neq}y \,|\, X{=}x) - P(Y{\neq}y \,|\, X{=}x)}{R(Y{\neq}y \,|\, X{=}x)} = 1 - U_{Y\neq y \triangleleft X=x}. \end{aligned}$$

With the help of this equation, $\hat{U}$ can be rewritten in the form

$$\hat{U}_{Y=y \lhd X=x} = \max\{1 - U_{Y \neq y \lhd X=x}\,, 0\} - \max\{1 - U_{Y=y \lhd X=x}\,, 0\}.$$

The inverse normalized preferences are defined analogously.

The preferences U are symmetrical for potential mates, given by the actual mating frequencies $P(Y{=}y)$ and $P(X{=}x)$. In contrast, $\hat{U}$ is symmetrical for these potential mates only when it is negative. Otherwise, both $\hat{U}_{Y=y \lhd X=x}$ and $\hat{U}_{X=x \lhd Y=y}$ are positive, and symmetry is realized only for $P(Y{=}y) = P(X{=}x)$.

In the following examples of application, in which population structure is considered solely for the reproducing members of populations, frequencies of potential mates will be assumed to equal their actual mating frequencies.

Metapopulations: reproductive isolation and coherence

In particular, if only membership of a group is relevant for maternal parents X and paternal parents Y, and if groups are denoted by z, then $\hat{U}_{Y=z \lhd X=z}$ specifies the degree to which maternal parents from group z prefer to mate with paternal parents from the same group (homodemic mating preferences). Positive values of $\hat{U}_{Y=z \lhd X=z}$ state that, on the average, maternal parents from group z prefer matings with paternal parents from their own group over matings with paternal parents from outside the group. This is reversed for negative values of $\hat{U}_{Y=z \lhd X=z}$. Analogous statements hold for the preferences $\hat{U}_{X=z \lhd Y=z}$ of paternal for maternal parents. In other words, group z is reproductively isolated from other groups and thus forms a subpopulation, to the degree to which both homodemic preferences $\hat{U}_{Y=z \lhd X=z}$ and $\hat{U}_{X=z \lhd Y=z}$ are positive.

Hence, if the $\hat{U}_{Y=z \lhd X=z}$ and $\hat{U}_{X=z \lhd Y=z}$ are strictly positive for all groups z, a clear tendency towards formation of a metapopulation, with respect to the chosen group structure, can be stated. Looking at metapopulation structures from the opposite point of view, i.e. in terms of degrees to which matings are performed among groups, the reproductive coherence or gene flow among groups is to be quantified. The relevant measures are the heterodemic preferences $\hat{U}_{Y \neq z \lhd X=z}$ and $\hat{U}_{X \neq z \lhd Y=z}$. The exact complementarity of both views is reflected by the mathematical identity $\hat{U}_{Y=y \lhd X=x} = -\hat{U}_{Y \neq y \lhd X=x}$. Positive values for $\hat{U}_{Y \neq z \lhd X=z}$ and $\hat{U}_{X \neq z \lhd Y=z}$ therefore indicate the absence of metapopulation structures for the chosen grouping criterion.

To arrive at a single *measure of metapopulation structure*, the differences between the sexes must again be taken into account. This suggests that we should distinguish between the average homodemic mating preferences of maternal for paternal parents, i.e. $\sum_z \hat{U}_{Y=z \lhd X=z} \cdot P(X{=}z)$, and of paternal for maternal parents, i.e. $\sum_z \hat{U}_{X=z \lhd Y=z} \cdot P(Y{=}z)$.

The special case of self-fertilization

When group structure is broken down to the level of the individual, so that each individual is considered a group of its own, the above homodemic mating preferences correspond to self-fertilization. It is thus meaningful in this case to talk about a measure of genealogical "self-preference" (the computation of U follows in this case the rules stated in the previous chapter, in connection with effects of spatial distance on mating preferences). The asymmetry in the measure $\hat{U}$ takes care of the possibility that self-fertilization has a female and a male component which may differ. Thus $\hat{U}_{Y=z \lhd X=z} > \hat{U}_{X=z \lhd Y=z} > 0$ states that parent z self-fertilizes more than at random and it does so to a larger degree as a maternal than as a paternal parent. At an extreme, when parent z is completely self-incompatible or produces for other reasons no offspring by self-fertilization, one obtains $\hat{U}_{Y=z \lhd X=z} = \hat{U}_{X=z \lhd Y=z} = -1$, which correctly reflects the implied complete rejection of homodemic matings.

Note that negative values of self-preference do not exclude self-fertilization but rather state that the reference individual mates more frequently with other individuals than with itself. In other words, despite a partial self-incompatibility, estimates of the proportion of offspring resulting from self-fertilization may be positive. This is an important mating system characteristic that does not show up in the common estimates of proportions of self-fertilization. Estimates of large proportions of self-fertilization may also be the result of random fertilization combined with low reproductively effective population sizes. In this case, high proportions of self-fertilization would be due to non-intact performance of the second fundamental function of mating systems. The third function is not directly affected by low reproductively effective population sizes. The inbreeding depression that might show up in the next generation is not due to the mating preferences but rather to the loss of adaptational capacity due to the drift effects associated with the small reproductive population size. Misinterpretations of this kind are, in fact, ruled out by consideration of the values of self-preference, since these would be zero because of the random fusion of the gametes.

Indirect (model-dependent) estimation

In the introductory remarks, it was indicated that the legitimacy of indirect estimates rests on tests of the validity of the underlying model. Models for which no experimental methods of testing are available are of limited practical relevance. This applies particularly to models which are used for the estimation of characteristics of real systems. Thus, if the mixed mating model is used in the estimation of the proportion of self-fertilization, the estimate is arbitrary if no test as to the validity of the model was performed. Estimates obtained from models which did not pass a statistical test are without substance. The estimation of amounts of gene flow among populations with the help of F_{ST} was mentioned in this context.

It is, of course, common practice to discuss unexpected indirect estimates in terms of the validity of the underlying model. Occasionally, sensitivity analyses are performed to rule out the possibility of substantial misinterpretation of data on the basis of the model. Most frequently, however, detailed arguments refer to more tractable problems, such as those arising from sample variance of estimators or from estimation algorithms connected with the respective model. Problems of testing the validity of the applied model are then of lesser concern (a more recent account of this situation is given e.g., by Ivey & Wyatt, 1999). This is at odds with the system analytic requirement for the joint consideration of model-dependent parameter estimation and model testing (see e.g. the paper of Gregorius, 1999, which demonstrates this requirement for the non-equilibrium and equilibrium version of the mixed mating model; a more comprehensive account of the system analytic approach is found in Gregorius 1998).

Intactness of mating mechanisms

The utilization of model-dependent methods of estimating mating system characteristics enforces integration of elements of parameter estimation and intactness analysis in the same model. This should be considered in the model design. For example, if the classical mixed mating model (selfing and random cross-fertilization) is used for the estimation of individual selfing rates, and if these estimates are used in a model of the effects of neighborhood mating including selfing, the assumptions on the form of cross-fertilization are contradictory between the two models. Inferences as to the intactness of the neighborhood mating mechanisms can thus not be consistently made, even if the neighborhood mating model would include the information on mating success required in an analysis if intactness.

Similar caution has to be taken with experiments in which the parentage (usually paternity) analysis is restricted to offspring of special origin such as seed from a few seed trees. Since intactness inferences of mating system mechanisms are to be based on samples that are representative of the population's offspring (zygote) production, the seed tree sample should represent the population, and the seed samples should represent each tree's seed production. Particularly the representativity of seed samples of the individual trees' seed production is frequently difficult to realize. This problem is more easily settled if seed is sampled after dispersal. Yet, this is achieved at the expense of reliable information on the maternal parent. However, as long as affiliation of seeds to population is unambiguous and representativity is guaranteed, the difference between sampling strategies affects the analysis only through statistical precision.

Acknowledgment – The author is indebted to Bernd Degen for providing the means to present and discuss this paper at the 2000 Symposium: "Modelling and Experimental Research on Genetic Processes in Tropical and Temperate Forests". The thoughtful suggestions and linguistic advice of Peter Smouse helped to improve the argumentation of the paper.

References

Adams WT, DS Birkes 1991. Estimating mating patterns in forest tree populations. Pp. 157-172 in Fineschi et al. 1991

Fineschi S, ME Malvolti, F Cannata, HH Hattemer (eds) 1991. Biochemical Markers in the Population Genetics of Forest Trees. SPB Academic Publishing bv

Gregorius H-R 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Math. Biosci. 41: 253-271

Gregorius H-R (ed) 1985. Population Genetics in Forestry. Lecture Notes in Biomathematics 60, Springer-Verlag

Gregorius H-R 1989. Characterization and Analysis of Mating Systems. Ekopan Verlag, Witzenhausen, Germany. http://webdoc.sub.gwdg.de/ebook/y/2001/gregorius/matesys.pdf

Gregorius H-R 1998. The system analytical approach to the study of hypotheses. http://webdoc.gwdg.de/ebook/y/1998/gregorius/sysana.pdf

Gregorius H-R 1999. Qualified testing of models and character estimation: the system analytic approach. Grüne Reihe der Sektion Biometrie und Informatik des Deutschen Verbandes Forstlicher Forschungsanstalten, 11. Jahrestagung: 253-260

Gregorius H-R, B Degen 1994. Estimating the extent of natural selection in seedlings from different *Fagus sylvatica* (L.) populations: Application of new measures. J. of Heredity 85: 183-190

Gregorius H-R, M Ziehe, MD Ross 1987. Selection caused by self-fertilization. I. Four measures of self fertilization and their effects on fitness. Theor. Pop. Biol. 31: 91-115

Haldane JBS 1954. The measurement of natural selection. Proc. 9^{th} Int. Congr. Genetics: 480-487

Hamrick JL, A Schnabel 1985. Understanding the genetic structure of plant populations: Some old problems and a new approach. Pp. 50-70 in Gregorius 1985.

Ivey CT, R Wyatt 1999. Family outcrossing rates and neighborhood floral density in natural populations of swamp milkweed (*Asclepias incarnata*): potential statistical artefacts. Theor. Appl. Genet. 98: 1063-1071

Lloyd DG 1979. Some reproductive factors affecting the selection of self-fertilization in plants. Am. Nat. 113: 67-79

Schoen DJ, AHD Brown 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA 88: 4494-4497

Steiner W, H-R Gregorius 1997. Reinforcement of genetic coherence: a single-locus model. BioSystems 43: 137-144

Wallace AR 1889. Darwinism – An Exposition of the Theory of Natural Selection With Some of Its Applications. Macmillan, London

Weir BS 1996. Genetic Data Analysis II. Sinauer Associates, Sunderland

Whitlock MC, DE McCauley 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. Heredity 82: 117-125

Willson MF 1994. Sexual selection in plants: Perspective and overview. Am. Nat. 144: 13-39

Ziehe M, H-R Gregorius 1988. Selection caused by self-fertilization. II. Ecological constraints on selfing advantage. J. evol. Biol. 1: 233-253

Appendix I: Selection load

A concept of viability selection load generalizing the classical approach of Haldane (1954) was developed by Gregorius & Degen (1994). An analogous generalization of this concept to include reproduction can be obtained by considering all adult individuals to have the same maximum ability (capacity) to reproduce. Reproductive output (number of offspring) can be measured in terms of zygotes, gametes, or successful gametes. Whenever the actual number of offspring of an individual falls below its reproductive capacity, impairment of the reproductive output due to environmental challenges or mating system effects can be stated. Making use of the notation

$$\begin{aligned} n_i &:= \text{number of } i\text{-type adults} \\ f_i &:= \text{number of offspring of } i\text{-type adults} \\ n &:= \textstyle\sum_i n_i \\ f &:= \textstyle\sum_i f_i \\ p_i &:= \frac{n_i}{n} \\ p_i' &:= \frac{f_i}{f} \\ c &:= \text{individual capacity of reproduction} \\ & \text{in terms of numbers of offspring,} \end{aligned}$$

the absolute reproductive impairment of the i-th type amounts to $n_i \cdot c - f_i$ offspring. By definition, c is always sufficiently large to assure non-negativity of this difference for all types. Hence, $c \geq f_i/n_i$ or all i, so that

$$c \geq \max_i \frac{f_i}{n_i} = \frac{f}{n} \cdot \max_i \frac{p_i'}{p_i}$$

It follows that the minimum reproductive capacity $c = c^*$, which must have been realized per individual to produce the observed numbers of offspring of the various types, equals $c^* = \max_i(f_i/n_i)$. The absolute total reproductive impairment across all types sums to $\sum_i (n_i \cdot c - f_i) = n \cdot c - f$, so that the relative impairment yields a fraction $\frac{n \cdot c - f}{n \cdot c}$ by which the total reproductive output is reduced relative to the total reproductive capacity. Since c^* is the minimum individual reproductive capacity that must have been realized to explain the differences in reproductive output between the types, c^* must be substituted for c to obtain the minimum relative reproductive impairment. One therefore arrives at

$$L_R := \frac{n \cdot c^* - f}{n \cdot c^*} = 1 - \frac{f}{n \cdot c^*} = 1 - \frac{1}{\max_i \frac{p_i'}{p_i}} = 1 - \min_i \frac{p_i}{p_i'}$$

as the minimum reduction in total reproductive output required to arrive at the realized differences in reproductive output between types. The fraction L_R is thus meaningfully addressed as the *reproduction selection load* to distinguish it from the viability selection load.

Comparison with the viability selection load L as stated in Gregorius & Degen (1994) shows that L is formally identical to the present reproduction load L_R, if the number of offspring of a type is measured by the number of individuals of this type remaining after selection. Hence, the above concept of reproduction selection load can indeed be generalized to include all vegetative and generative stages. One only has to refer the n_i's and f_i's to any two successive developmental stages, the first of which being a vegetative (including zygotic) and the second a vegetative or generative stage. This justifies denotation of L_R simply as the selection load.