BMC Evolutionary Biology

Research article

Reinforcement of genetic coherence in a two-locus model

Hans-Rolf Gregorius* and Wilfried Steiner

Address: Institut fur Forstgenetik und Forstpflanzenzüchtung Universität Göttingen, Büsgenweg 2, D-37077 Göttingen, Germany

E-mail: Hans-Rolf Gregorius* - gregorius@gwdg.de; Wilfried Steiner - nfv-abtc@t-online.de

*Corresponding author

Published: 7 August 2001

BMC Evolutionary Biology 2001, **1**:2

Received: 22 May 2001

Accepted: 7 August 2001

This article is available from: http://www.biomedcentral.com/1471-2148/1/2

Abstract

Background: In order to maintain populations as units of reproduction and thus enable anagenetic evolution, genetic factors must exist which prevent continuing reproductive separation or enhance reproductive contact. This evolutionary principle is called genetic coherence and it marks the often ignored counterpart of cladistic evolution. Possibilities of the evolution of genetic coherence are studied with the help of a two-locus model with two alleles at each locus. The locus at which viability selection takes place is also the one that controls the fusion of gametes. The second locus acts on the first by modifying the control of the fusion probabilities. It thus acts as a mating modifier whereas the first locus plays the role of the object of selection and mating. Genetic coherence is enhanced by modifications which confer higher probabilities of fusion to heterotypic gametic combinations (resulting in heterozygous zygotes) at the object locus.

Results: It is shown that mutants at the mating modifier locus, which increase heterotypic fusions but do not lower the homotpyic fusions relative to the resident allele at the object locus, generally replace the resident allele. Since heterozygote advantage at the object locus is a necessary condition for this result to hold true, reinforcement of genetic coherence can be claimed for this case. If the homotypic fusions are lowered, complex situations may arise which may favor or disfavor the mutant depending on initial frequencies and recombination rates. To allow for a generalized analysis including alternative models of genetic coherence as well as the estimation of its degrees in real populations, an operational concept for the measurement of this degree is developed. The resulting index is applied to the interpretation of data from crossing experiments in *Alnus* species designed to detect incompatibility relations.

Introduction

Anagenetic (phyletic) and cladogenetic evolution can be basically distinguished by the fact that during the former genetic variation is transformed within a single population without losing the reproductive contact between the genetic variants, while in the latter genetic variation is distributed to reproductively separated populations. In other words, phyletic evolution has the capacity to maintain or strengthen "genetic coherence" among the genetic variants. This coherence is lost as a consequence of reproductive separation during cladistic processes. The necessity to consider these complementary processes as of equal significance in evolutionary reasoning was recognized, for example, by [1] and becomes already evident in the running title "Can speciation be prevented?" of this paper. Contrary to common concepts, the title suggests the existence of persistently acting forces of genetic

disjunction that have to be counteracted in order to maintain genetically variable reproductive communities.

In its probably most obvious form the separation-coherence dualism becomes relevant in hybrid zones, where genetic separation between the hybridizing populations is apparent for some traits but not for others. Hybrid zones may therefore be considered as a more or less stable balance between speciation and coherence. A concise review of the mechanisms that could be responsible for this situation is provided e.g. by [2]. As far as mating relations are considered as potential mechanisms they are confined to reinforcement of prezygotic isolation through hybrid disadvantage. The problems with experimental verification of reinforcement as well as with its consistent modeling are pointed out in a recent review by [3]. In view of these complications it might not be surprising that, according to these reviews, possibilities of reinforcing the internal reproductive coherence as mechanisms which enable populations to maintain their genetic integrity do not seem to have attracted any attention.

Yet, as the present authors demonstrated in a series of papers [4–7], the apparent evolutionary complementarity of reproductive separation and coherence in fact has fundamental genetic substance and can even be derived from Wallace's early theory of speciation based on the evolutionary reinforcement of reproductive isolation in cases of hybrid disadvantage [8], p.l75ff, called the Wallace effect in [9]. Replacing "hybrid" by "heterozygote", inversion of Wallace's idea allows to reformulate Felsenstein's running title as "does heterozygote advantage reinforce genetic coherence?" Herewith, reinforcement of genetic coherence is to be understood as the replacement of extant genetic types by mutants that increase mating preferences among different genetic types (increase heterotypic mating preferences; for the concept of mating preferences see [10]).

For a single-locus, three-allele model involving pleiotropic effects on survival and mating traits, the present authors demonstrated that Wallace's extended concept of the reinforcement of mating preferences holds true. Thus, for pleiotropic gene action on viability and mating preferences, heterozygote disadvantage reinforces the evolution of homotypic mating preferences (avoidance of heterotypic matings), and heterozygote advantage reinforces heterotypic mating preferences.

When viewed in the framework of genetic load it turns out that Wallace's theory can actually be extended to imply that reinforcement of the respective mating preferences simply reduces the genetic load without sacrificing adaptively relevant genetic variation by reducing the formation of unfit genotypes. In this way, adaptability is maintained at lower costs for adaptedness and population integrity and persistence are thus enhanced. Yet, this is so far confirmed only for pleiotropic gene action [7]. For non-pleiotropic gene action, which requires at least two gene loci, confirmation of this principle is limited to speciation [6]. Its counterpart, genetic coherence, still awaits modeling and analysis. The present paper is devoted to this topic. The model design will follow the two-locus principle argued by [4], where one locus modifies the mating relations realized at a second locus, and where this second locus is also subject to selection.

Since the above concept of genetic coherence embraces a continuum of mating (and gene flow) relations which extend from complete avoidance of heterotypic matings (completion of speciation) to exclusively heterotypic mating (complete reproductive coherence), the present paper will also be concerned with the development of an index which quantifies the different degrees of reproductive coherence. This index is intended to aid in recognizing evidence for genetic coherence in population genetic data. Its range of application will be demonstrated for an analysis of data from crossing experiments in *Alnus* species which were designed to detect incompatibility relations.

The Model

Description of the model

A model with two biallelic loci A and *B* is considered in which viability selection (in the diplophase) is restricted to the genotypes at the *B*-locus. The *B*-locus is also involved in the formation of zygotes in that the alleles present at this locus in the encountering gametes of different sex specificity determine their probabilities of fusion. The role of the *A*-locus consists solely in modifying these fusion probabilities according to the alleles carried by the encountering gametes at this locus. No selection occurs at the *A*-locus. As suggested by [4], the *A*- and *B*-locus will be referred to as the mating modifier and the object locus, respectively. Gametes of different sex specificity are assumed to encounter at random. The loci are linked with recombination rate *r* (the relevant notational details are compiled in Table 1 and illustrated in Figure 1). Generations are assumed to be separated and population size is effectively infinite.

Mating modification via gamete fusion

As detailed in Table 1, the probability that a pair of gameteS G_i^k and G_j^l fuses to a zygote after an encounter is described by the probability f_{ij}^{kl}. In the case that a pair of gametes does not fuse after an encounter, both gametes are assumed to be incapable of further reproductive activity. This establishes a mating system described for

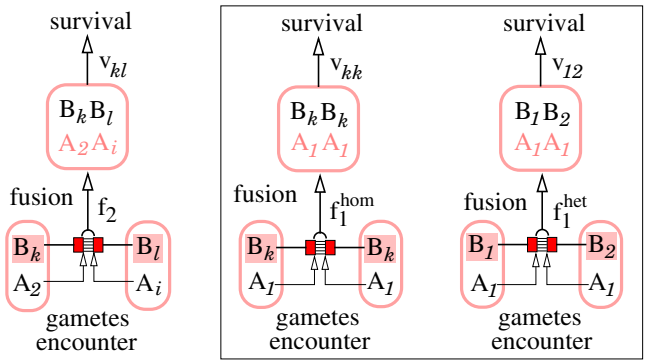

Figure 1
Schematic representation of the model determinants of reproduction and survival: (1) gametes encounter at random, (2) gamete fusion is determined by the *B*-locus (the object locus of mating), and fusion probabilities are modified by the *A*-locus (the mating modifier locus), (3) survival is determined by the *B*-locus (the object locus of survival) only; the *B*-locus thus is object of both gamete fusion and survival. The framed part refers to differential fusion modification with respect to homotypic and heterotypic encounters at the object locus *B*.

plants as "selective fertilization with pollen- and ovule-elimination" (see e.g. [11–13]). Such mating systems affect both the combination of gametes into zygotes and the mating success. Differential mating success is implied by fact that, even though all gametes of one sex may have the same chance to encounter a gamete of the other sex, they are differentially successful since they fail to fuse when encountering an "incompatible" gamete. The mating success of an individual thus depends on the frequencies of gametic types produced in the population which are "compatible" with the gametic types produced by this individual.

Since, basically, genetic coherence refers to mating relations among different in comparison to like genetic types, and since these relations are defined for the object locus *B*, fusion probabilities must reflect this fact. In particular, if the probabilities of fusion among gametes are the same for all allelic combinations at the *B*-locus, random fusion can be stated for this locus. This is in fact the situation of the absence of any genetic coherence, and it will be assumed to be realized for the wild type allele A_2 at the mating modifier locus. Since recessivity is generally believed to be the most likely tpye of gene action for newly arising mutants, dominance of the wild type over the mutant A_1 will be assumed in the effect on gamete fusion. Thus the probabilities of fusion are the same for all allelic combinations at the *B*-locus if at least one of the two encountering gametes carries A_2. This probability will be denoted by f_2, and it is characterized by $f_2 := f_{2i}^{kl}$ for all, *k*, *l*.

In a homotypic encounter for the mutant mating modifier allele A_1 it is assumed that the probability of fusion of the gametes is the same for the two homotypic associations at the object locus *B*, i.e. $f_{11}^{11} = f_{11}^{22} =: f_1^{hom}$. This is a reasonable assumption in view of the fact that

Table 1: Notation General remarks

A	modifying locus with two alleles A_1 and A_2
B	object locus with two alleles B_1 and B_2
G_i^k	two-locus gamete (A_i B_k) with allele *i* at the *A*- and allele *k* at the *B*-locus
G_{ij}^{kl}	zygote or any other diploid genotype originating from fusion of gametes (A_i B_k) and (A_j B_l). Note that because of unordered genotypes $G_{ij}^{kl} = G_{ji}^{lk}$ but because of linkage $G_{ij}^{kl} \neq G_{ij}^{lk}$
a_i, b_k	relative frequencies of alleles A_i, B_k among adults
g_i^k	relative frequency of gamete G_i^k in the gametic production
g_{ij}^{kl}	relative frequency of genotype G_{ij}^{kl} among adults
a_{ij}, b_{kl}	relative frequency of genotype A_i A_j and B_k B_l, respectively, among adults
r	recombination frequency ($0 \le r \le 1/2$)
f_{ij}^{kl}	probability of fusion when the gametes G_i^k and G_j^l encounter. *f*-values are assumed to differ in only three ways: f_2 := probability of fusion in the presence of allele A_2 in at least one of the two encountering gametes, f_{2i}^{kl} for all *i*, *k*, *l*; f_1^{hom} := probability of fusion when A_1-carriers with equal *B*-alleles encounter (homotypic mating), $f_{11}^{11} = f_{11}^{22} = f_1^{hom}$; f_1^{het} := fusion probability among A_1-carriers with different *B*-alleles (heterotypic mating), $f_{11}^{12} = f_{11}^{21} = f_1^{het}$;
v_{kl}	viabilities at the *B*-locus
s_{ij}^{kl}	combined selection value of fusion probability and viability selection, defined as $s_{ij}^{kl} := f_{ij}^{kl} \cdot v_{kl}$

The designation of gene loci, alleles, gametes, and genotypes is by uppercase letters, relative frequencies and probabilities are indicated by lowercase letters. The indexing for two-locus-types is done such that *A*- and *B*-locus alleles appear as subscripts and superscripts, respectively. The indices appearing one upon the other indicate alleles located on the same gamete prior to or at fusion. A prime indicates next generation frequencies.

changes in genetic coherence show primarily in alterations of the fusion probabilities for heterotypic combinations at the object locus. These probabilities will be distinguished by the notation $f_1^{het} := f_{11}^{12} = f_{11}^{21}$. Figure 1 may again serve as an illustration of these details.

Transition equations

Gamete formation

Starting from a situation which is characterized by a frequency distribution of the 10 possible two-locus-genotypes G_{ij}^{kl}, the gamete formation results after regular recombination at a rate r and for $i \neq j$, $k \neq l$ as

$$g_i^k = g_{ii}^{kk} + \tfrac{1}{2}g_{ii}^{kl} + \tfrac{1}{2}g_{ij}^{kk} + \tfrac{1}{2}(1-r)g_{ij}^{kl} + \tfrac{1}{2}r g_{ij}^{lk}$$

Zygote formation

The assumption of random encounter of gametes of different sex yields pairs of (yet non-fused) gametes in the following relative frequencies (with $i \neq j$, $k \neq l$):

$$\begin{aligned} G_i^k \times G_i^k &: \quad (g_i^k)^2 \\ G_i^k \times G_i^l &: \quad 2g_i^k g_i^l \\ G_i^k \times G_j^k &: \quad 2g_i^k g_j^k \\ G_i^k \times G_j^l &: \quad 2g_i^k g_j^l \end{aligned}$$

The fusion probabilities decide about final zygote formation so that the new generation starts with zygotic frequencies

$$\begin{aligned} g_{11}^{kk*} &= (g_1^k)^2 f_1^{hom}/\bar{f} \\ g_{22}^{kk*} &= (g_2^k)^2 f_2/\bar{f} \\ g_{11}^{12*} &= 2g_1^1 g_1^2 f_1^{het}/\bar{f} \\ g_{22}^{12*} &= 2g_2^1 g_2^2 f_2/\bar{f} \\ g_{12}^{kk*} &= 2g_1^k g_2^k f_2/\bar{f} \\ g_{12}^{12*} &= 2g_1^1 g_2^2 f_2/\bar{f} \\ g_{12}^{21*} &= 2g_1^2 g_2^1 f_2/\bar{f}, \end{aligned}$$

with the average probability of fusion $\bar{f} = \sum_k \big((g_1^k)^2 f_1^{hom} + (g_2^k)^2 f_2 + 2g_1^k g_2^k f_2\big) + 2\big(g_1^1 g_1^2 f_1^{het} + g_2^1 g_2^2 f_2 + g_1^1 g_2^2 f_2 + g_1^2 g_2^1 f_2\big)$ as normalization factor. The asterisk indicates that these frequencies refer to the phase after gametic fusion and before viability selection.

Viability selection

The zygotic genotypic frequencies resulting from random encounter of gametes and subsequent formation of zygotes according to the probabilities of fusion is now subjected to viability selection at the *B*-locus so that the genotypic frequencies among the adults of the new generation become: $g_{ij}^{kl\prime} = g_{ij}^{kl*} v_{kl}/\bar{v}$ for all i, j, k, l with the average viability $\bar{v} = \sum_{k\leq l} v_{kl} \sum_{i\leq j} g_{ij}^{kl*} + v_{12} g_{12}^{21*}$ as normalization factor. Insertion of the above genotypic frequencies after fusion from equations (2) into equations (3) leads to

$$\begin{aligned} g_{11}^{kk\prime} &= (g_1^k)^2 f_1^{hom} v_{kk}/(\bar{v}\bar{f}) \\ g_{22}^{kk\prime} &= (g_2^k)^2 f_2 v_{kk}/(\bar{v}\bar{f}) \\ g_{11}^{12\prime} &= 2g_1^1 g_1^2 f_1^{het} v_{12}/(\bar{v}\bar{f}) \\ g_{22}^{12\prime} &= 2g_2^1 g_2^2 f_2 v_{12}/(\bar{v}\bar{f}) \\ g_{12}^{kk\prime} &= 2g_1^k g_2^k f_2 v_{kk}/(\bar{v}\bar{f}) \\ g_{12}^{12\prime} &= 2g_1^1 g_2^2 f_2 v_{12}/(\bar{v}\bar{f}) \\ g_{12}^{21\prime} &= 2g_1^2 g_2^1 f_2 v_{12}/(\bar{v}\bar{f}), \end{aligned}$$

where the explication of $\bar{v}\bar{f}$ yields $\bar{v}\bar{f} = \sum_k \big((g_1^k)^2 f_1^{hom} + (g_2^k)^2 f_2 + 2g_1^k g_2^k f_2\big) v_{kk} + 2\big(g_1^1 g_1^2 f_1^{het} + g_2^1 g_2^2 f_2 + g_1^1 g_2^2 f_2 + g_1^2 g_2^1 f_2\big) v_{12}$.

This supplies us with the transition equations between successive adult stages. Intermediate stages are given by equations (1) and (2).

The following anaylses are organized along steps of increasing complexity of interaction between the two loci, starting with consistent effects of each locus and ending

with numerical analyses of complex effects suggested by the preceding analytical results.

Analysis of the dynamics – analytical characterizations

The focus is on the conditions under which the allele A_1 may become established and eventually fixed in the population. Since this allele is considered to increase the probability of fusion for gametes which carry different alleles at the *B*-locus (the object locus), its dynamics decides on the evolution of increased genetic coherence at the *B*-locus.

It is in fact possible to greatly simplify the analysis by considering that equations (4), when combined with equation (1), result in a representation of the transition equations which is mathematically equivalent to the classical two-locus model of viability selection and random mating (random fusion of gametes). In this representation, each encounter leads to fusion of the gametes. The resulting zygote has a probability of survival which is identical to the probability of fusion of its constituent gametes. This first viability selection phase affects both loci *A* and *B* jointly. A second phase is characterized by viability selection restricted to the *B*-locus, so that overall one obtains two-locus genotypic viabilities of the form $s_{ij}^{kl} = f_{ij}^{kl} \cdot v_{kl}$. The *s*- values will be termed "combined selection value" in the following. This allows us to apply results known from analyses of the classical model.

Following the general concept suggested by [14] for the analysis of two-locus polymorphisms, we may start with observing for each genotype at one locus the ranking of the three genotypes at the other locus. For example, if $s_{ij}^{11} < s_{ij}^{12} > s_{ij}^{22}$ for all *i,j*, then we have complete conditional overdominance in viability and thus a protected (stable) polymorphism at the *B*-locus irrespective of the recombination rate. Similarly, $s_{22}^{kl} \leq s_{12}^{kl} \leq s_{11}^{kl}$ for all *k, l* implies fixation of A_1 for all recombination rates if the sign of inequality is strict in at least one case. In terms of the restrictions of the present model, these cases can be more easily pictured by looking at the rows and columns in the arrangement of the two-locus viabilities presented in Table 2.

From inspection of the columns it becomes clear that $f_1^{het} \geq f_1^{hom}$ guarantees protectedness of the *B*-locus polymorphism if there is overdominance at this locus, i.e. $v_{12} > \max\{v_{11}, v_{22}\}$. Recall that $f_1^{het} \geq f_1^{hom}$ is required in connection with questions of the evolution of increased coherence. Yet, even $f_1^{het} < f_1^{hom}$ need not destabilize the *B*-locus polymorphism provided $f_1^{het}/f_1^{hom} \geq \max\{v_{11}, v_{22}\}/v_{12}$.

Table 2: Combined selection values

s_{ij}^{kl}	A_1A_1	A_1A_2	A_2A_2
B_1B_1	f_1^{hom}	f_2v_{11}	f_2v_{11}
B_1B_2	f_1^{het}	f_2v_{12}	f_2v_{12}
B_2B_2	31	f_2v_{22}	f_2v_{22}

Looking at the rows one notes that allele A_1 would become fixed if $\min\{f_1^{het}, f_1^{hom}\} \geq f_2$ and if not both f_1-values are equal to f_2. The special case $f_1^{het} > f_1^{hom} = f_2$ is included in this condition, and it is of direct relevance to the evolution of coherence, since it states that the mutant A_1 increases the fusion probabilities only among gametes differing at the *B*-locus. When realized together with $v_{12} \geq \max\{v_{11}, v_{22}\}$, this implies that the *B*-locus polymorphism is protected and, by fixation of A_1, increased genetic coherence becomes established. The expectation that overdominance reinforces genetic coherence is so far confirmed.

If one aims at more general results it must be taken into account, that increased coherence cannot evolve via substitution of A_2 by A_1 unless the *B*-locus polymorphism is maintained for fixation of A_1, i.e. unless $f_1^{het} \cdot v_{12} > f_1^{hom} \cdot \max\{v_{11}, v_{22}\}$, which is equivalent to $f_1^{het}/f_1^{hom} > \max\{v_{11}, v_{22}\}/v_{12}$. In principle, this necessary condition includes situations other than overdominance at the *B*-locus $(v_{12} > \max\{v_{11}, v_{22}\})$. Such situations, however, allow only for transient *B*-locus polymorphisms prior to the advent of the A_1 mutant. Depending on the allele frequencies at the *B*-locus, it is conceivable that the A_1 mutant replaces the wild type and by this may stabilize the *B*-locus polymorphism. However, since this situation is only locally stable it is of limited interest for the evolution of increased genetic coherence.

Another case not ruled out by the necessary condition is the possibility that f_2 is located between f_1^{het} and f_1^{hom}. If, in addition to the symmetry in homotypic fusion probabilities $(f_{11}^{11} = f_{11}^{22} = f_1^{hom})$ there exists symmetric overdominance in the viabilities at the *B*-locus $(v_{11} = v_{22} < v_{12})$; the polymorphism at this locus will be protected and the frequencies of the alleles B_1 and B_2 will ultimately become equal. The reason is to be seen in the fact that both *B*-homozygotes have equal average *s* values which stay below the pertaining average $B_1 B_2$-viability for each single *A*-genotype. As a consequence, initially present stochastic associations between the two loci will decay and the average viability of $A_1 A_1$ will ultimately approach $\frac{1}{2}(f_1^{hom} v_{hom} + f_1^{het} v_{12})$, where $v_{hom} = v_{11} = v_{22}$. By the same reasoning the average viabilities of the other two *A*-genotypes both will approach $\frac{1}{2} f_2 (v_{hom} + v_{12})$. On the basis of these average viabilities it can be stated that A_1 will or will not replace A_2 according to whether the difference $\frac{1}{2}(f_1^{hom} v_{hom} + f_1^{het} v_{12}) - \frac{1}{2} f_2 (v_{hom} + v_{12})$ and thus $v_{12}(f_1^{het} - f_2) - v_{hom}(f_2 - f_1^{hom})$ is positive or negative.

It is straightforward to show that these analytical results still hold, if the mutant mating modifier A_1 is assumed to be dominant over the wild type A_2. All one has to do is replace f_2 in the $A_1 A_2$-column of Table 2 by f_1^{het} and f_1^{hom}, repectively, and then follow the above steps of analysis.

However, simple predictions of the dynamics for asymmetric overdominance in viability at the *B*-locus ($v_{11} \neq v_{22}$) are not possible. Since the analytical treatment of these cases is generally quite intricate, the following considerations will resort to numerical analyses of selected scenarios.

Analysis of the dynamics – numerical studies

According to the results of the above analytic considerations it remains to study situations of asymmetric overdominance in viabilities at the *B*-locus ($v_{11} \neq v_{22}$, $v_{kk} < v_{12}$) in combination with effects of A_1 that increase heterotypic fusions and decrease homotypic fusions $(f_1^{het} > f_2 > f_1^{hom})$. This will be done with the help of five scenarios concerning the specification of viability, fusion, and recombination parameters and initial conditions for genotypic frequencies. The scenarios Sl to S5 are compiled in Table 3. Since our major concern is the study of the conditions for establishment and ultimate fixation of increased coherence, all scenarios are characterized by initial conditions for genotypic frequencies, in which the mutant A_1 is represented at low frequency and frequencies at the *B*-locus are close to equilibrium.

Table 3: Scenarios Numerical scenarios for overdominance in viability at the B-locus

no:	S1	S2*	S3*	S4*	S5*
f_1^{hom}	0:3	0:1			
f_1^{het}	0:8				
f_2	0:5				
v_{11}	0:4				
v_{12}	0:5				
v_{22}	0:3		0:1	0:1	0:1
r	0:3			0:01	0:01
g_{11}^{11}	0				
g_{11}^{12}	0				
g_{11}^{22}	0				
g_{12}^{11}	0				
g_{12}^{12}	0:006				0:001
g_{12}^{21}	0				
g_{12}^{22}	0				
g_{22}^{11}	0:043				
g_{22}^{12}	0:618				0:623
g_{22}^{22}	0:333				
A_1	fix.	elim.	elim.	fix.	elim.

*Scenarios are identical to S1 except of the values given in the respective columns. elim.: elimination of A_1; fix.: fixation of A_1

A typical example is represented by scenario S1 with a relatively strong coherence effect of A_1 as expressed in the *f*-ranking $(f_1^{het} = 0.8, f_2 = 0.5, f_1^{hom} = 0.3)$. The over-dominance reinforces the increase of A_1 as a factor favoring heterotypic fusions. During the first 1500 generations A_1 increases steadily but rarely exceeds 1%. The genetic structure at locus *B* during this phase is therefore only little modified by A, b_1 running rapidly

very close to 0.6666 (the equilibrium frequency in the absence of different *f*-values). The pronounced increase of A_1 between generations 2200 and 2400 towards fixation entails a decrease of b_1 to 0.5254.

(i) coherence by increasing f_1^{het} vs. decreasing f_1^{hom}

The coherence effect results from higher fusion probability for heterotypic encounters on the one hand and from reduced fusion probability for homotypic encounters on the other. While the first effect promotes increase of A_1 and a simultaneous increase of the population fitness, the second effect enables a rapid change of the frequency structure at locus *B* but at the same time reduces the population fitness and especially the fitness of $A_1 A_1$ types, thus reducing the chances for establishment and fixation of coherence. If scenario S1 is modified by reducing f_1^{hom} to 0.1 (S2), the overall disadvantage of A_1 results in a continuous decrease of A_1 with the exception of a small increase during the first 10 generations. Whether the combination of increased $f_1^{het} > f_2$ and decreased $f_1^{hom} < f_2$ results in higher or lower fitness of A_1 compared to A_2 depends on the two-locus genotypic structure. As allelic and genotypic frequencies change during the dynamics, the direction of the dynamics may change as is the case in scenarios Sl and S2, for example.

(ii) Degree of viability asymmetry

If S1 is modified by reducing v_{22} to 0.1 (other parameters, especially the *f*-values, remaining unchanged, see S3), this increased disadvantage for homotypic encounters could be expected to favor the evolution of coherence. The numerical results, however, are in contrast to such intuitive expectation: After an initial increase during the first 12 generations, A_1 decreases and will be lost. The reason must be seen in the fact, that the asymmetric *v*-values lead to an equilibrium at *B* with a significantly higher frequency of B_1 and therefore to an increased frequency of homotypic encounters. In this situation, the fitness reducing f_1^{hom} cannot be compensated by the fitness increasing f_1^{het} as is the case in S1.

(iii) influence of recombination rate r

Starting with S3 as an example for failure of establishment of coherence, and reducing *r*, it can be seen that for recombination rates of 0.05 or higher the increase of A_1 is prevented despite an initial increase (from 0.3% to 0.6% for $r = 0.05$, for example). For $r = 0.04$ or smaller, the initial increase is also followed by a phase of decrease, the dynamics' direction, however, is reversed once more at about generation 110 (for $r = 0.04$) or generation 160 (for $r = 0.01$, see S4) leading to fixation of A_1 in about 2000 generations.

(iv) influence of starting frequencies

Using scenario S4 as a reference for successful establishment and fixation of coherence, a reduction of the starting frequency a_1 from 0.3% to 0.05% (S5) results in the failure of establishment despite an initial increase until generation 22 ($a_1 = 0.11\%$).

Measuring Genetic Coherence

To allow for a generalized analysis including alternative models of genetic coherence as well as the estimation of its degrees realized in actual populations, it is desirable to provide an operational concept for the measurement of this degree. For this purpose recall that genetic coherence and genetic separation are opposite evolutionary concepts which refer to the tendency for each allele to preferentially occur in association with other allelic types (heterozygosity) or with its own type (homozygosity) in diploid genotypes. Preferential association of an allele with its own type indicates isolation against and thus separation from other alleles. Such an allele thus contributes to the reproductive fragmentation of a population. Hence, an index *C* of genetic coherence should be specified for each allele, and it should attain its lowest value if it occurs only in association with its own type (complete isolation), while associations only with other types should determine its largest value. The borderline between coherence and separation is drawn by the situation where an allele is associated with its own type exactly in proportion to its occurrence in the population. For the *B*-locus with allele frequencies b_k and homozygote frequencies b_{kk} this implies that the index *C* reaches its lower bound for $b_{kk} = b_k$ and its upper bound for $b_{kk} = 0$. The borderline, where the allele shows no preferential associations with its own nor with other types is reached at $b_{kk} = b_k^2$.

A conceptually consistent construction of such an index is achieved by making use of the above-mentioned concept of mating preferences in the form introduced by [10]. The mating preference $U_{l \triangleleft k}$ of type *k* for type *l* is there defined by the ratio $P_{l \triangleleft k}/R_{l \triangleleft k}$ where $P_{l \triangleleft k}$ and $R_{l \triangleleft k}$ are the actual and potential frequencies, respectively, of type *l* mates among all mates of type *k*. The preferences $U_{l \triangleleft k}$ are unbounded and are equal to 1 in the absence of any preferences of type *k* for type *l* (indifference, random mating). Yet, given the distribution of potential mates, $U_{l \triangleleft k}$ is bounded from above by $1/R_{l \triangleleft k}$

(since $P_{l \triangleleft k} \leq 1$), and this bound characterizes the situation of complete preference of type *k* for type *l*. Along the same reasoning, $U_{l \triangleleft k} = 0$ characterizes complete rejection of *l*-type mates by *k*-types. To arrive at a measure of mating preference that varies symmetrically around the situation of indifference and extends over the range from complete rejection to complete preference, it is desirable to normalize *U* accordingly. The normalized version $\hat{U}$ should ideally assume values of -1, 0, and +1 for complete rejection, indifference, and complete preference, respectively. This is realized by

$$\hat{U}_{l \triangleleft k} := \begin{cases} U_{l \triangleleft k} - 1 & \text{if } U_{l \triangleleft k} \leq 1 \\ \dfrac{U_{l \triangleleft k} - 1}{R^{-1}_{l \triangleleft k} - 1} & \text{if } U_{l \triangleleft k} > 1 \end{cases}$$

Substitution of "allele" for "type" and considering the formation of a zygote as a mating event, the conditional mating frequencies $P_{l \triangleleft k}$ are given by $^1/_2 b_{lk}/b_k$ for $l \neq k$ and by b_{kk}/b_k for $l = k$. Since coherence is to be measured only among the gametes which entered the formation of zygotes, the "potential mates" are the same for all alleles and equal their frequencies among the zygotes, i.e. $R_{l \triangleleft k} = b_l$. Hence, $U_{l \triangleleft k} = \frac{1}{2} b_{lk}/(b_l \cdot b_k)$ for $l \neq k$ and $U_{k \triangleleft k} = b_{kk}/b_k^2$. By the above reasoning, coherence is characterized by heterotypic preferences and thus by preferential associations of one allele with alleles other than its own type. If the subscript *l* collectively denotes all alleles other than *k*, one obtains for the heterotypic preferences $U_{l \triangleleft k} = (b_k - b_{kk})/((1-b_k) \cdot b_k)$ with $R_{l \triangleleft k} = 1 - b_k$. Taking account of the desired normalization, this suggests to define an index C_k of coherence for the *k*-th allele by the heterotypic preference $\hat{U}_{l \triangleleft k}$ and thus by

$$C_k := \begin{cases} \dfrac{b_k^2 - b_{kk}}{b_k(1-b_k)} & \text{if} \quad b_k^2 \leq b_{kk} \ (\leq b_k) \\ \dfrac{b_k^2 - b_{kk}}{b_k^2} & \text{if} \quad b_k^2 \geq b_{kk} \end{cases}$$

In closed form, C_k can be written as

$$C_k = \frac{b_k^2 - b_{kk} + (1 - 2b_k) \cdot |b_k^2 - b_{kk}|}{2(1-b_k) b_k^2}$$

If the *k*-th allele does not occur in heterozygotes, it is reproductively completely isolated from other alleles as is characteristic of a biological species. In this case $b_k = b_{kk}$ and thus $C_k = -1$. At the other extreme, for gametophytic incompatibility systems each allele occurs only in heterozygotes, so that $b_{kk} = 0$ and therefore $C_k = 1$. For each such allele complete genetic coherence can be stated.

For two alleles the two C_k's strongly depend on each other, since then $b_1^2 - b_{11} = b_2^2 - b_{22} = \frac{1}{2} b_{12} - b_1 b_2$. Hence, C_1 and C_2 always have the same sign and, in addition, for $b_k^2 - b_{kk} \leq 0$ (i.e. $C_k \leq 0$) (i.e. $C_k \leq 0$) both *C*-values are even identical, i.e. $C_1 = C_2$. On the other hand, if $b_k^2 - b_{kk} > 0$, the relation between the two allelic coherence indices are determined by the two allele frequencies, since then $C_1/C_2 = (b_2/b_1)^2$. The less frequent allele shows in this case the larger coherence. In other words, for homozygote excess (relative to Hardy-Weinberg proportions) both alleles contribute equally to the population's genetic coherence, while for heterozygote excess the less frequent allele contributes more to genetic coherence than the predominant allele.

Taking the average over the C_k's, i.e. $\bar{C} = b_1 \cdot C_1 + b_2 \cdot C_2$, yields $\bar{C} = b_{12}/(2b_1 b_2) - 1$ irrespective of the sign of the C_k's. This relates $\bar{C}$ to Wright's fixation index *F* by $\bar{C}$. Hence, *F* allows for an interpretation that is usually not directly associated with concepts of genetic coherence or separation/speciation. The model-independence of the concept underlying the C_k's thus enlarges the scope of application of *F* to the interpretation of data on genotypic frequencies obtained for stages close to the zygotic stage. The lower and upper bounds of $\bar{C}$ for given allele frequencies at the *B*-locus are realized for $b_{12} = 0$, which yields $\bar{C}$, and for $^1/_2 b_{12} = \min\{b_1, b_2\}$, which yields $\bar{C} = \max\{b_1, b_2\}^{-1} - 1$ (for further details concerning boundaries of heterozygosity and *F* see e.g. the book of [15]).

In numerous computer runs of the present model it turned out that $\bar{C}$ increases with the frequency of the allele A_1 that enhances heterotypic fusions. This did not generally hold for the dynamics of each of the individual

C_k-values. Since the C_k-values become positive as A_1 approaches fixation, this is possible according to the above analysis. If A_1 entails increased heterotypic fusions without lowering homotypic fusions, both *C*-values increase continuously with the establishment of A_1.

For the case of multiple alleles (> 2), *C*-based analyses of genotypic structures can be extended to more complex problems by forming groups of alleles which are considered equivalent in some defined sense. In the case of four alleles B_1,...,B_4 for example, B_1 and B_2 can be considered as equivalent and to form a group B_x, say, with allele frequency $b_x = b_1 + b_2$ and "homozygote" frequency $b_{xx} = b_{11} + b_{12} + b_{22}$. The coherence measure of this composite allele B_x is then well defined by C_x, and degrees of reproductive isolation from or coherence with other alleles or groups of alleles can be analyzed. It should, however, be noted that the average $\bar{C} = \sum_k C_k \cdot b_k$ need no longer equal -*F*, since for multiple alleles $F = 1 - (1 - \sum_k b_{kk})/(1 - \sum_k b_k^2)$ interpretation of -*F* as average genetic coherence is thus limited to two alleles.

An application

The applicability of the coherence indices C_k covers a range, which exceeds that of genotypic structures of populations. In the context of the present model involving selective fertilization, outcomes of controlled crosses are of particular interest, since they allow direct observation of fusion probabilities at the gametic level. Such crosses were performed by the present authors in a project concerned with the detection of mating incompatibility relations in *Alnus* species. Crosses between parents with the same heterozygote genotype at various isoenzyme gene loci yielded genotypic frequencies among their seed which differed significantly from the hypothesis of regular segregation and random fusion of the gametes. Two examples are provided by samples of 39 : 73 : 15 for A_1A_1 : A_1A_2 : A_2A_2 at the SKDH-A locus in one cross, and 12 : 8 : 5 for B_2B_2 : B_2B_4 : B_4B_4 at the 6PGDH-B locus in another cross.

The *C*-values for the SKDH locus are C_1 = 0.131, C_2 = 0.282 and $\bar{C}$ = 0.192, while for the 6PGDH locus these values become $C_2 = C_4 = \bar{C}$ = -0.306. Note, that *C*-values are based on successful gametes only, so that they are not affected by segregation distortion but rather reflect solely effects of fusion preferences. Ignoring sampling effects, these observations suggest strongly opposing tendencies for the two loci, with homotypic fusion preferences at the 6PGDH locus and heterotypic preferences at the SKDH locus. This need, of course, not indicate the existence of opposing forces acting functionally at the two enzyme loci. Structural associations of the enzyme loci via chromosomal coupling with functionally effective loci in the genetic background may as well serve for an explanation. These functional loci, however, must in both crosses be assumed to be heterozygous in at least one crossing partner of each of the two crosses to explain the observations. It is also clear that in the case of the 6PGDH locus the two alleles must have been in coupling phase with the preferentially fusing alleles at the functional locus. In the same way, the two SKDH alleles must be assumed to be in repulsion phase with the preferentially fusing alleles at the functional locus.

In any case, this observation of strongly opposing effects at different loci can be expected to extend to the whole population only in the absence of noticeable stochastic associations between the loci. The reason is that an allele with a strongly positive de-gree of genetic coherence and a distinctly positive association with an allele at another locus prohibits strongly negative genetic coherence for this other allele. Consequently, (sympatric or parapatric) speciation can be initiated only at loci which show no associations with loci that exhibit high degrees of genetic coherence, and each of the speciating subpopulations inherits the genetic coherence relations of the base population.

Conclusions

Two factors are considered in the two-locus model presented in this paper: the mating system (with fusion probabilities at the *B*-locus depending on the allelic composition at the A-locus) and classical viability selection (at the *B*-locus). It turned out that the particular specification of the mating system allows for an evolutionarily equivalent interpretation of the model in terms of a two-locus viability selection model with random mating, the combined selection values s_{ij}^{kl} resulting as the product of the fusion probability and the one-locus viability: $s_{ij}^{kl} = f_{ij}^{kl} \cdot v_{kl}$ (see Table 2). Coherence may be promoted in two ways, by increasing f_1^{het} or by decreasing f_1^{hom}. In the first way the fitness of Ai carriers is increased, and in the second way it is decreased.

A necessary prerequisite for the evolution of coherence is a stable polymorphism at the object locus *B*. Since the selection coefficients in our model are not frequency-dependent, overdominance is required to ensure the persistence of polymorphism. Prior to the appearance of any mating modifier, "simple" overdominance ($v_{11} < v_{12} > v_{22}$) is sufficient for a stable *B*-polymorphism. With

mating modifiers present in the population, complete conditional overdominance at the *B*-locus $(s_{22}^{kl} < s_{12}^{kl} > s_{11}^{kl})$ is required for a stable polymorphism, and this is equivalent to $f_1^{het}/f_1^{hom} \geq \max\{v_{11}, v_{22}\}/v_{12} < 1$. Given this condition, the mutant A_1 will replace the resident allele A_2 if $\{f_1^{het}, f_1^{hom}\} \geq f_2$ and $\max\{f_1^{het}, f_1^{hom}\} > f_2$. Since the latter inequalities include the case $f_1^{het} > f_1^{hom} = f_2$ thus increased coherence, the expectation that overdominance reinforces genetic coherence is confirmed so far. Conclusively, reversion of these inequalities, which characterizes decreased coherence for the mutant, prohibits its establishment.

The situation $f_1^{het} \geq f_2 \geq f_1^{hom}$ is ambiguous with respect to its effect on genetic coherence. Yet, the analytical results obtained for symmetric viability ($v_{11} = v_{22} < v_{12}$) show that A_1 will become fixed or eliminated according to whether $v_{12}(f_1^{het} - f_2) - v_{hom}(f_2 - f_1^{hom})$ is positive or negative. Thus, if f_1^{het} exceeds f_2 by a sufficient amount, the evolution of increased coherence is again reinforced. For intermediate f_2 and non-symmetric v_{kk}'s, however, numerical studies demonstrated more complex dynamics. Examples are presented where a change only in the initial frequency of A_1 or in the recombination rate turns the dynamics from fixation to elimination of A_1 or vice versa.

Application of the conceptually generalized measures of coherence confirmed these results. Therefore, in essence the inverse of Wallace's principle holds for the present model: heterozygote superiority not only prevents the evolution of reproductive separation of subpopulations but even reinforces the evolution of increased genetic coherence.

A single-locus model of reinforcement of genetic coherence previously suggested by the present authors ([7]) allowed for a complete analysis covering a much wider range of mating systems. For this model it was shown that overdominance in viability generally implies the replacement of a resident allele by a mutant conferring higher heterotypic mating preferences, while heterozygote disadvantage generally promotes the evolution of higher homotypic mating preferences. This clear dualism does not seem to exist in the present two-locus model despite its more detailed mating system. Fusion probabilities, viability parameters and recombination rates interact in more complex ways. Even if the selective effects of the mating system are separated from its purely combinational effects, as was possible in the one-locus model, recombination apparently introduces dynamical forces which become dominant for compensating forms of selection and combination.

Acknowledgement

This work was supported by grant Gr435/17-1 of the Deutsche Forschungsgemeinschaft. The authors are thankful to two anonymous reviewers for their constructive comments.

References

1. Felsenstein J: **Skepticism towards Santa Rosalia, or why are there so few kinds of animals?** *Evolution* 1981, **35**:124-138
2. Barton NH, Hewitt GM: **Adaptation, speciation and hybrid zones.** *Nature* 1989, **341**:497-503
3. Noor MAF: **Reinforcement and other consequences of sympatry.** *Heredity* 2000, **83**:503-508
4. Gregorius H-R: **Population genetic keys to speciation.** *Göttingen Research Notes in Forest Genetics* 1992, **13**:1-19
5. Gregorius H-R: **A single-locus model of speciation.** *Acta Biotheoretica* 1992, **40**:313-319
6. Gregorius H-R: **A two-locus model of speciation.** *Journal of theoretical Biology* 1992, **154**:391-398
7. Steiner W, Gregorius H-R: **Reinforcement of genetic coherence: a single-locus model.** *BioSystems* 1997, **43**:137-144
8. Wallace AR: *Darwinism – An Exposition of the Theory of Natural Selection With Some of Its Applications. Macmillan, London, New York,* 1889
9. Grant V: **The selective origin of incompatibility barriers in the plant genus Gilia.** *Amer. Nat* 1966, **100**:99-118
10. Gregorius H-R: *Characterization and Analysis of Mating Systems. Ekopan Verlag, Witzenhausen, Germany* 1989 [http://webdoc.sub.gwdg.de/ebook/y/2001/gregorius/matesys.pdf]
11. Schwemmie und Mitarbeiter J: **Genetische und cytologische Untersuchungen an Eu-Oenotheren.** *Zeitschr. f. indukt. Abstamm.- u. Vererb.-Lehre* 1938, **75**:358-800
12. Finney DJ: **The equilibrium of a self-incompatible polymorphic species.** *Genetica* 1952, **26**:33-64
13. Haustein E: **Die selektive Befruchtung.** *in: W Ruhland (ed.) Sexualität – Fortpflanzung – Generationswechsel. Handbuch der Pflanzenphysiologie, Band 18, Springer: Berlin, Heidelberg, New York,* 1967479-505
14. Gregorius H-R: **Establishment of allelic two-locus polymorphisms.** *J. Math. Biol* 1991, **30**:185-197
15. Weir BS: *Genetic Data Analysis II. Sinauer Associates,* 1996