SELECTION AT WORK IN SELF-INCOMPATIBLE *ARABIDOPSIS LYRATA* II.
SPATIAL DISTRIBUTION OF S-HAPLOTYPES IN ICELAND

Mikkel H. Schierup, Jesper S. Bechsgaard and Freddy B. Christiansen

Department of Ecology and Genetics, Institute of Biology, University of Aarhus, 8000 Aarhus C, Denmark
Running head: Spatial distribution of S-haplotypes

Keywords: Arabidopsis lyrata, sporophytic self-incompatibility, spatial distribution, island model, dominance

Corresponding author:

Mikkel Heide Schierup
Bioinformatics Research Center (BiRC)
Department of Ecology and Genetics
University of Aarhus
Hoegh Guldbergsgade 10, building 1090
DK-8000 Aarhus C
Denmark

Phone: +45 89423231
Fax: +45 89423077
Email: mheide@birc.au.dk
http://www.birc.au.dk/~mheide
We survey the distribution of haplotypes at the self-incompatibility locus of *Arabidopsis lyrata* (Brassicaceae) at 12 locations spread over the species’ natural distribution in Iceland. Previous investigations of the system have identified 34 functionally different S-haplotypes maintained by frequency-dependent selection and arranged them into four classes of dominance in their phenotypic expression. Based on this model of dominance and the island model of population subdivision, we compare the distribution of S-haplotypes with that expected from population genetic theory. We observe 18 different S-haplotypes, recessive haplotypes being more common than dominant ones, and dominant ones being shared by fewer populations. As expected, differentiation, although significant, is very low at the S-locus even over distances of up to 300 km. The frequency of the most recessive haplotype is slightly larger than expected for a panmictic population, but consistent with a subdivided population with the observed differentiation. Frequencies in nature reflect effects of segregation distortion previously observed in controlled crosses. The dynamics of the S-locus variation are, however, well represented by a 12-island model and our simplified model of dominance interactions.
INTRODUCTION

Most homomorphic self-incompatibility (SI) systems are determined by a single multi-allelic locus (the S locus), composed of at least two genes. In the sporophytic self-incompatibility (SSI) system of Brassicaceae, the S-receptor kinase (SRK) is responsible for the determination of female specificity, and the S cysteine-rich protein (SCR) is responsible for the determination of male specificity (SCHOPFER et al. 1999; TAKASAKI et al. 2000). In a panmictic population with gametophytic SI the dynamics were modelled quite accurately by Sewall Wright already in 1939 (WRIGHT 1939). Gametophytic self-incompatibility systems are simple, in that incompatibility is determined by matching of a haploid pollen with a diploid stigma expressing both its haplotypes codominantly. In a panmictic population all haplotypes are expected to have a frequency drawn from the same unimodal distribution. Sporophytic self-incompatibility systems are found within several families, in particular Brassicaceae (BATEMAN 1954). They are more complex to model because pollen phenotype derives from the genotype of the pollen parent, and dominance may occur in both stigma and pollen (BATEMAN 1952). In simple sex-symmetric dominance models where selection on the male and female sides is equally strong, phenotypes are expected to have the same mean frequency (although with different distributions) leading to a higher expected frequency of relatively recessive haplotypes because they display their phenotype less often (BATEMAN 1952; CHARLESWORTH 1988; CHARLESWORTH et al. 2000; COPE 1962; IMRIE et al. 1972; SAMPSON 1967; SCHIERUP et al. 1997). In cases of more complicated dominance relations among the haplotypes, the homogeneity of phenotype frequencies fails (SCHIERUP et al. 2006). Recessive haplotypes, however, still show relatively higher frequencies, as found in surveys of sporophytic systems (MEHLENBACHER 1997; SCHIERUP et al. 2006).

Population structure is common in plant species, stimulating theoretical and empirical studies on the effect of population structure on the distribution of systems under balancing selection. These theoretical studies show that the strong frequency-dependent selection causes very different dynamics for SI haplotypes as compared to a neutral gene (MUIRHEAD 2001; SCHIERUP 1998b; SCHIERUP et al. 2000a;
SCHIERUP et al. (2000b). Haplotypes at very low frequencies show high fitness compared to those close to their equilibrium frequency. Invasion of a haplotype absent in a subpopulation therefore has a high probability of success. In particular, it is higher than that of neutral alleles, leading to the prediction that population differentiation is lower at the S locus than at neutral loci. The number of S haplotypes in a subdivided population is not expected to be an increasing function of the degree of isolation as for neutral genes. Actually, for most reasonable migration rates a subdivided population maintains fewer haplotypes than a panmictic population with the same number of individuals (MUIRHEAD 2001; SCHIERUP 1998b). These conclusions apply to both gametophytic and sporophytic SI. The study by GLEMIN et al. (2005), using an antibody assay in Brassica insularis, is consistent with these predictions. Furthermore, population structure is expected to affect haplotypes at different dominance levels differently. Assuming a sex-symmetric model, SCHIERUP et al. (2000b) found that more recessive haplotypes have higher average frequencies in a subdivided population than in a panmictic one, mainly because they are present in a larger proportion of subpopulations. More dominant haplotypes were found to be lost locally more often and therefore have a higher effective migration rate and corresponding shorter coalescence time for copies of the same haplotype.

In studies of the SI system of Arabidopsis lyrata (Brassicaceae) we have previously characterized 34 different SRK haplotypes of the SRK orthologue (BECHSGAARD et al. 2006; CHARLESWORTH et al. 2000) and developed a PCR based typing system for identification of S-haplotypes at the SRK gene (MABLE et al. 2005; MABLE et al. 2003; SCHIERUP et al. 2006). This typing system was used to study the mating patterns within a single Icelandic population (SCHIERUP et al. 2006). Here we describe the frequency of S haplotypes in 12 populations across Iceland. The dominance relations among the haplotypes are relatively well known (PRIGODA et al. 2005). In a subdivided population we therefore are able to develop predictions on equilibrium frequencies and the dynamics of haplotypes as a function of their dominance level. The observations in nature are compared to deterministic theory based on either the number of haplotypes observed locally or in the global population, and to results from computer simulations explicitly modelling the observed population structure. We conclude that
the dynamics of the system are well captured by our dominance model and the ensuing intrinsic balancing selection, in particular if we allow for limited differentiation among the populations.

MATERIALS AND METHODS

Arabidopsis lyrata is one of the most common plant species in Iceland. It grows mainly on lava fields, disturbed habitat, along roads, and throughout the highlands. Leaf material and seeds from 12 different locations was collected during a field trip in August 2001. The geographical positions of the sampled populations are shown in Figure 1 with more details in Table 1. At each sampling site, 30-100 individuals were sampled.

S haplotypes were determined for 30-50 individuals per population using the same protocol as in Schierup et al. (2006). The overall strategy is first to use two sets of general primers (13SEQ1F and 13SEQ2F forward with SLGR reverse), and then to restriction digest the resulting PCR fragments using two four-cutter enzymes, AluI and MspI. The resulting electrophoretic banding pattern suggests the haplotypes. These are subsequently confirmed with haplotype-specific primer sets. In order to identify the missing haplotypes, we subjected the individuals with only one initially identified haplotype to further analyses using two general forward primers (13F1 and 13SEQ3F). If this did not result in the discovery of a second haplotype, we tested primers specific for haplotypes known not give very strong signals with general primers (haplotype ALSRK36, ALSRK37 and ALSRK38, with the primer sets ALSRK36F1 – ALSRK36R1, ALSRK37F1 – ALSRK37R1 and Aly13-38 – 38-R1, respectively). If this failed to identify a second haplotype, we consider the individual to be homozygote for its haplotype. All haplotypes except for ALSRK38 have previously been reported and linkage to the S-locus established for all but ALSRK38 and ALSRK27 (Bechsgaard et al. 2006; Mable et al. 2005; Mable et al. 2003; Schierup et al. 2006; Schierup et al. 2001). The sequence of ALSRK38 has been submitted to GenBank (accession number XXXX). The sequences of all specific primer sets can be found in supplementary Table S1.
**Statistical analysis:** A phylogenetic tree of the S-haplotypes was reconstructed based on 650 bp of the S-domain sequenced in all the haplotypes identified in the present study using Aly3 and Aly7 as outgroups (CHARLESWORTH *et al.* 2003). The minimum-evolution criterion with the HKY substitution model was used and analysis was performed in Mega 3 (KUMAR *et al.* 2004). The phylogeny was used to infer dominance class membership of haplotypes where this has not been established (haplotypes ALSRK27 and ALSRK38).

F-statistics and Mantel test for isolation by distance (correlation between matrices of $F_{ST}$ values and geographical distances) were calculated using Fstat version 2.9.3.2 (GOUDET 1995).

Mate availability within each population was estimated by calculating the mate availability for individual plants using the current model of dominance relationship between the haplotypes in stigma and pollen (SCHIERUP *et al.* 2006). The mate availability of an individual is the proportion of the population that it can fertilise, thus we assume that selection occurs through male function only, i.e. there is no variation in female fecundity (VEKEMANS *et al.* 1998). This assumption reflects the fact that full seed set is typically observed in the field (SCHIERUP *et al.* 2006).

The mean and variance in frequency of haplotypes across populations were quantified, and the homogeneity of the variance for different haplotypes was analyzed.

**Numerical evaluations of haplotype frequencies:** The dynamics of a sporophytic SI system with variable number of haplotypes at four dominance levels were investigated through iterations of a deterministic model and through stochastic computer simulations. Dominance was assumed complete between dominance levels in both sexes, with codominance within dominance levels. The model is thus a combination of the SSIcond and SSIdom models of SCHIERUP *et al.* (1997), and it therefore assumes
random mating and male sexual selection. Investigations including selection on the female side (fecundity selection, (VEKEMANS et al. 1998)) gave very similar results, but they are not reported.

The expected equilibrium frequency of haplotypes in a population was determined as in SCHIERUP et al. (2006). The deterministic recurrence equations were iterated assuming symmetry in genotypic frequencies with respect to the haplotypes within any of the dominance levels, and can be considered a special case of the framework recently published by BILLIARD et al. (2007). When the genotypic frequencies of the haplotypes within a particular dominance level are equal at the start of the iteration, they remain so throughout the iterations. Two types of iterations were run. One where each subpopulation was run separately using the observed number of haplotypes for each dominance level in that subpopulation. The other investigated the total population using the total number of haplotypes observed in the entire population. Stability of equilibria was checked by perturbations followed by iterations omitting the symmetry assumption. The Mathematica notebook iterating the recurrence equations was written by F.B.C. and can be downloaded through http://www.daimi.au.dk/~mheide/.

The stochastic simulations are based on individual plants fertilized from a local pollen pool formed from equal contributions from all local plants and modified by immigration. A 12-island model of population structure was implemented with migration at the same rate between all islands as in Wright’s island model. Only pollen were allowed to migrate, modifying the local pollen pool by replacing a given fraction by an equal mixture of pollen from the population. Different combinations of population sizes and number of subpopulations were tested, keeping the total population size between 1200 and 2400. All subpopulations were assumed to be of the same size—sizes of 100 or 200 were used in accord with the population structure observed in previous studies (SCHIERUP 1998a; SCHIERUP et al. 2006). A Poisson-distributed random number of already existing haplotypes was introduced into the population each generation to avoid loss of haplotypes. The probability of introduction of a given haplotype was proportional to its expected frequency under the deterministic equilibrium, and the process can be thought of as immigrants from a larger population, e.g. the remaining part of Iceland.
The rate of introduction was set to 0.5%, which turned out to be sufficient to avoid loss of haplotypes in the simulations. Simulations were initiated with a set of genotypes drawn at random from the equilibrium frequency distribution determined by the deterministic iterations, and they were continued until a selection-drift-mutation steady state was attained (at least 5,000 generations). At this time a sample of 30 individuals was picked from each population, and frequencies of haplotypes were calculated as well as \( F_{ST} \) for the set of sampled populations. The sample size was chosen in order to mimic the sampling strategy used in the empirical survey. The migration rate was adjusted by trial and error for each set of parameters until \( F_{ST} \) values were obtained similar to the observed \( F_{ST} = 2\% \). The local immigration fraction was thus set at 0.12 when the population size is 100 and at 0.08 with 200 individuals in the local population (Other simulations with different number of individuals in the populations but the level of differentiation at 2% gave comparable results). The structure of the natural population is presumably not reflected by this simple model, but its genetic structure is well approximated. Of course, the correspondence between the simulated structure and that of the collection of sampled natural populations is not of a quality that allows the simulation results to be the foundation upon which formulation of quantitative hypotheses suited for statistical analyses may be built. The simulation results are thus qualitative guidelines for the interpretation of patterns in natural observations.

To probe the effect of yet finer population structure we simulated a two-dimensional finite stepping stone model with immigration at equal rates from the four neighbouring subpopulations. The 225 subpopulations of size 10 are arranged on a torus to avoid edge effects. Total immigration was set at 0.3 in each subpopulation, and that rate yielded \( F_{ST} \) values in the range 2-4%. Otherwise, details of the simulation were as in the 12-island model. When evaluating population differentiation, the torus was partitioned into 25 areas of 90 individuals by merging 9 adjacent subpopulations (a 3 by 3 section). \( F_{ST} \) was then calculated between these 25 areas. The simulation programs were written by M.H.S. in Python. The source code with limited documentation is also available through

www.daimi.au.dk/~mheide/.
RESULTS

Eighteen out of 38 different S-haplotypes known in the species were identified (Table 2), and these fall into four different dominance classes, with 1 recessive haplotype, 3 of the least dominant haplotypes, 3 of the dominant haplotypes and 11 of the most dominant haplotypes (Figure 2). Throughout this study we assume an absolute dominance relation among these four classes with codominance within, and these dominance relations are equal in the sexes.

Two haplotypes were identified in $271/405=67\%$ of individuals, and only one haplotype was observed in the remaining 33\% of individuals. All individuals showed at least one haplotype. Single-haplotype individuals are assumed to be homozygotes for this haplotype, even though we cannot rule out that they are heterozygotes for an unidentified haplotype (see an extended discussion of this below).

Observed frequencies of haplotypes in each population are shown in Figure 3. More dominant haplotypes are generally less common, and the most recessive haplotype is the most common in all populations. Since haplotypes from the same dominance level are expected to have the same frequencies, we may consider the total frequency of haplotypes at each level. These are in Figure 4 compared to deterministic calculations that predict the total frequency of haplotypes at each dominance level in each population, on the assumption that we have sampled all the haplotypes present in each population.

**Population structure and haplotype frequency dynamics:** Average differentiation of haplotype frequencies among populations as measured by $F_{ST}$ is only 2\%, but highly significant (P<0.001). We investigated the nature of this differentiation in more detail. Table 3 shows an analysis of the homogeneity of the dominance-class counts among populations, concluding clear evidence of population structure most likely due to variation in all class frequencies. Figure 5 shows $F_{ST}$’s between pairs of populations as a function of distance between the two populations, indicating no correlation...
between F_{ST} and distance (Mantel-test, P>0.6). This lack of isolation by distance suggests the use of Wright’s 12-island model in simulations to elucidate the differentiation among sampled populations. Absence of some of the haplotypes from local populations was also seen in the simulations (Table 4), so the heterogeneous pattern seen within dominance classes in Figure 3 need not be ascribed to limited sample sizes. This is true for the most dominant class, in particular. Comparing observations to expectations in a panmictic population (Table 5), the most striking deviation is that the observed frequency of the most recessive haplotype is higher than predicted. Computer simulations of a population subdivided into 12 subpopulations with a similar level of geographical differentiation produce results comparable to those obtained, when treating the twelve populations as isolated populations with different numbers of haplotypes. Comparing the observed haplotype frequencies to those obtained from simulations of local populations, where only the locally observed haplotypes are present, the most notable change is that the observed frequencies of the most recessive class are closer to those expected—the differences are small, though. Inclusion of the observed level of genetic population structure in the theoretical predictions therefore improves the correspondence between simulations and observations.

Simulations were also used to evaluate whether the large fraction (33%) of putative homozygotes can be explained by very fine scaled population structure. If no substructure exists within populations, and all populations are weighted equally, one would expect 13.8% homozygotes from deterministic calculations. Simulations using the island model of migration yielding F_{ST}=0.02 produce homozygote frequencies between 18% and 25% (range over 50 replicates). The effect of finer local population structure were investigated by simulating a two-dimensional stepping stone model (see Material and Methods). For comparable levels of geographical differentiation the simulations yielded the frequency of homozygotes in the range 25-30% (observed in 50 replicates). Thus, substructure within local populations has the potential of significantly increasing homozygote frequencies through biparental inbreeding. The values found in this simulation are indeed closer to the observed value, but further
investigations are needed to resolve to what extent biparental inbreeding is the main cause of the observed excess of total homozygote frequency.

The 134 individuals observed with a single haplotype (and assumed to be homozygotes) are unequally distributed over dominance classes, with 81 homozygous for the recessive, 27 for the least dominant, 10 for the dominant and 16 for the most dominant class (supplementary data, Table S2).

**Mate availability:** Figure 6 shows the population wide frequencies of haplotypes as well as the observed average mate availability that each haplotype experiences in the total population. The very common recessive haplotype has the lowest mate availability. The mate availability in the local population decreases with the frequency of the haplotype. For the least dominant and dominant haplotypes this is reflected as a decrease in the average mate availability with the frequency of the haplotype, but for the most dominant haplotypes the relation between these averages is far from perfect reflecting deviations between haplotype frequencies and average genotype frequencies akin to those described as the Wahlund effect.

Mate availability in the 12 populations range between 82% and 92% with an average 88% chance that two individuals from the same population can mate. Compared to the expected mate availabilities calculated from the observed haplotype frequencies, all observed mate availabilities are low (Figure 7). Two individuals picked at random from the total population are compatible with a probability of 91%, again consistent with slight population differentiation. Different haplotypes have different mate availabilities (Figure 6), and the more dominant haplotypes generally have higher mate availabilities.
DISCUSSION

We report the presently most comprehensive typing of S-haplotypes in a subdivided natural population with the aim of testing whether the spatial distribution is concordant with our simplified models of dominance of S-haplotypes and population structure. We found that our model of the self-incompatibility system in *Arabidopsis lyrata* describes the dynamics of the system. The model assumption of a strict dominance hierarchy is, however, a simplification. Instances of reciprocal differences in crosses and dominance within dominance classes have been reported in *A. lyrata*, depending on the genomic background (MABLE et al. 2004; PRIGODA et al. 2005). Our deterministic calculations of expected frequencies of haplotypes as a function of their dominance level fit the observed patterns in the population individually. Thus, at the local level, the dynamic equilibrium of the dominance class frequencies appears to be reached quickly in accordance with the strong frequency-dependent selection. In this sense the dynamics of S-haplotypes in the local population are important for the dynamics of the SSI polymorphism in Iceland. However, frequencies of individual haplotypes within a class vary considerably due to random genetic drift within the local population and sampling variance due to local heterogeneity (SCHIERUP et al. 2006). Low but significant population structure makes the deterministic theory of a panmictic population insufficient for describing haplotype frequencies in the entire population. We thus resorted to computer simulations to illuminate the effects of the observed level of differentiation. Local absence of some of the most dominant haplotypes was seen in both natural and simulated populations. The higher expected effective migration rate of haplotypes with increasing dominance is therefore offset by their increased risk of local extinction by drift—in concordance with earlier results based on the simpler sex-symmetric SSIdom model with only one haplotype at each dominance level (SCHIERUP et al. 2000b). The observed frequencies of haplotypes at the various dominance levels are accurately predicted by simulations. The expected effect of population subdivision is therefore that the more dominant haplotypes have lower frequencies and higher mate availabilities, as observed in the present study.
We observe an excess of homozygotes over the theoretical expectation, assuming that all samples were taken from the same population. Part of this excess is due to local fluctuation in haplotype frequencies due to population structure. An island model of population structure, however, only predicts about 20% homozygotes, but 33% of the sampled individuals only show a single haplotype. The remaining excess of homozygotes can be ascribed to an unknown combination of more fine scaled population structure within sampling sites, inability to detect certain (unknown) haplotypes, incompletely known dominance relations, and a limited amount of selfing. Simulation of the first situation produced up to 30% homozygotes for reasonable parameter values. The individuals sampled in the 12 populations were typically taken from an area of 100x100m. This is a considerably larger scale than the 10x15 m area, where we previously detected fine scale genetic clustering of S haplotypes in a different Icelandic population (SCHIERUP et al. 2006).

Observation of homozygotes for haplotypes in the most dominant class cannot be explained by population structure but suggests one of three possibilities: 1) haplotypes exist that our typing does not recognize, 2) dominance exists within the most dominant class, or 3) selfing is possible. We are not able to distinguish between these alternatives with the present data. We believe that our methods detect all haplotypes that have previously been characterised in this species (34 in all), but we also find it plausible that haplotypes exist that cannot be detected at present. BECHSGAARD et al (2004) found evidence that at least one haplotype is not detected, and MABLE et al. (2003, 2004) found evidence for dominance within dominance classes. In addition, selfing has evolved repeatedly in A. lyrata in nature (MABLE et al. 2005; PRIGODA et al. 2005), and we found selfing in the greenhouse under stressful conditions (unpublished observation). Our results suggest that putative undetected haplotype(s) are relatively rare, and therefore most likely belong to the most dominant class. Undetected haplotypes are therefore unlikely to produce large quantitative effects on our results. Incomplete knowledge of dominance relations and their occasional breakdown can of course produce excess homozygotes, in particular if selfing is possible.
BECHSGAARD et al. (2004) found that the haplotypes ALSRK01 and ALSRK09 have segregation advantages in certain full-sib crosses. These two haplotypes have the smallest average mate availabilities in the present study. It is thus tempting to speculate that they indeed possess a segregation advantage in natural populations, and that this increases their frequencies to a level, where their mate availabilities decrease their male sexual fitness until their segregation advantage is balanced. However, our results show that segregation distortion in general is not sufficiently common to disturb the overall dynamics of the system away from that predicted by our simple model of dominance.

The lack of isolation by distance was not anticipated since *A. lyrata* continuously cover large parts of Iceland. Previous studies in Iceland have shown that pollen and seed dispersal in established populations is very limited (SCHIERUP et al. 2006). Isozyme studies have observed $F_{ST}>0.1$ among populations separated by less than 40 km, evidencing isolation by distance at this scale (BECHSGAARD et al. 2004; SCHIERUP 1998a; SCHIERUP and BECHSGAARD unpublished results). $F_{ST}$ values are, however, not directly comparable for isozymes and highly polymorphic genes under balancing selection. We can see two possible reasons for the low level of isolation by distance at the S locus. First, the overall level of differentiation of S-haplotypes is low due to their efficient introgression, causing isolation by distance undetectable on the investigated geographical scale. Second, the colonization of Iceland is very recent, although *A. lyrata*, like many other species of plants, might have survived the Weichselian glaciation at nunatak refugia in northern Iceland (RUNDGREN and INGOLFSSON 1999). The retreat of glaciers began about 10,000 years ago, and a significant fraction of the land is still glaciated (Figure 1). The present distribution is likely still to reflect a colonization process where the plants follow the receding rim of the ice. Typing of neutral molecular markers and S haplotypes in the same individuals is needed to distinguish effects of these scenarios.
ACKNOWLEDGMENTS

We thank Deborah Charlesworth, Barbara Mable, Xavier Vekemans and Vincent Castric for sharing unpublished material and for many valuable discussions. Camilla Håkansson is thanked for help in the lab. The study was supported by grants from the Carlsberg Foundation, the Danish Natural Sciences Research Council and the Danish Veterinary and Agricultural Research Council to MHS.

REFERENCES


GLEMIN, S., T. GAUDE, M. L. GUillaume, M. LOURMAS, I. OLIVIERI et al., 2005 Balancing selection in
the wild: testing population genetics theory of self-incompatibility in the rare species Brassica
insularis. Genetics.
GOUDet, J., 1995 FSTAT (Version 1.2): A computer program to calculate F-statistics. JOURNAL OF
HEREDITY 86: 485.
IMRIE, B. C., D. R. ROSS and C. J. KIRKMAN, 1972 Computer Simulation of a Sporophytic Self-
Incompatibility Breeding System. Australian Journal of Biological Sciences 25: 343-&.
KUMAR, S., K. TAMURa and M. NEI, 2004 MEGA3: Integrated software for Molecular Evolutionary
MABLE, B. K., J. BELAND and C. Di BERARDO, 2004 Inheritance and dominance of self-incompatibility
MABLE, B. K., A. V. ROBERTSON, S. DART, C. Di BERARDO and L. WITHAM, 2005 Breakdown of self-
incompatibility in the perennial Arabidopsis lyrata (Brassicaceae) and its genetic consequences.
MABLE, B. K., M. H. SCHIERUP and D. CHARLESWORTH, 2003 Estimating the number, frequency, and
dominance of S-alleles in a natural population of Arabidopsis lyrata (Brassicaceae) with
MEHLENBACHER, S. A., 1997 Revised dominance hierarchy for S-alleles in Corylus avellana L.
Theoretical And Applied Genetics 94: 360-366.
MUirHEAD, C. A., 2001 Consequences of population structure on genes under balancing selection.
PRIGODA, N. L., A. NASSUTH and B. K. MABLE, 2005 Phenotypic and genotypic expression of self-
incompatibility haplotypes in Arabidopsis lyrata suggests unique origin of alleles in different
RUNDGREN, M., and O. INGOLFSSON, 1999 Plant survival in Iceland during periods of glaciation?


SCHIERUP, M. H., and J. S. BECHSGAARD, unpublished results.


Table 1

Characteristics of locations sampled

<table>
<thead>
<tr>
<th>Population</th>
<th>GPS coordinates</th>
<th>Altitude</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21°57638W 64°27390N</td>
<td>62m</td>
<td>Roadside</td>
</tr>
<tr>
<td>2</td>
<td>21°03888W 65°08693N</td>
<td>28m</td>
<td>Field</td>
</tr>
<tr>
<td>3</td>
<td>21°23858W 65°28595N</td>
<td>32m</td>
<td>SW slope</td>
</tr>
<tr>
<td>5</td>
<td>20°57659W 65°26009N</td>
<td>68m</td>
<td>Roadside</td>
</tr>
<tr>
<td>6</td>
<td>17°53840W 65°51547N</td>
<td>74m</td>
<td>Roadside</td>
</tr>
<tr>
<td>8</td>
<td>17°28876W 65°33475N</td>
<td>200m</td>
<td>Roadside</td>
</tr>
<tr>
<td>11</td>
<td>17°49725W 65°03870N</td>
<td>833m</td>
<td>Stone desert</td>
</tr>
<tr>
<td>13</td>
<td>18°24895W 64°32785N</td>
<td>679m</td>
<td>Stone desert</td>
</tr>
<tr>
<td>14</td>
<td>18°49182W 64°22893N</td>
<td>593m</td>
<td>Stone desert</td>
</tr>
<tr>
<td>15</td>
<td>19°53538W 64°07191N</td>
<td>162m</td>
<td>Roadside</td>
</tr>
<tr>
<td>16</td>
<td>20°07851W 64°20742N</td>
<td>259m</td>
<td>Roadside</td>
</tr>
<tr>
<td>17</td>
<td>20°53756W 64°02238N</td>
<td>80m</td>
<td>NW slope</td>
</tr>
</tbody>
</table>
Table 2

Number of individuals sampled and their haplotypes

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of Individuals</th>
<th>Number of haplotypes classified</th>
<th>Number of homozygotes assumed</th>
<th>Haplotype sample size</th>
<th>Number of different haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>41</td>
<td>9</td>
<td>50</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>58</td>
<td>14</td>
<td>72</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>52</td>
<td>8</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>58</td>
<td>14</td>
<td>72</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>49</td>
<td>9</td>
<td>58</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>39</td>
<td>62</td>
<td>16</td>
<td>78</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>37</td>
<td>70</td>
<td>4</td>
<td>74</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>66</td>
<td>14</td>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>32</td>
<td>57</td>
<td>7</td>
<td>64</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>41</td>
<td>62</td>
<td>20</td>
<td>82</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>25</td>
<td>40</td>
<td>10</td>
<td>50</td>
<td>11</td>
</tr>
<tr>
<td>17</td>
<td>35</td>
<td>59</td>
<td>11</td>
<td>70</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>405</strong></td>
<td><strong>674</strong></td>
<td><strong>134</strong></td>
<td><strong>810</strong></td>
<td><strong>18</strong></td>
</tr>
</tbody>
</table>
Table 3. Results of chi-square tests of homogeneity of haplotype classes among populations. The recessive class is compared to the sum of the three dominant classes, the least dominant class is compared to the sum of the dominant and most dominant classes, and the dominant class is compared to the most dominant class. These tests are independent, and their values sum to the usual $4 \times 12$ homogeneity test.

<table>
<thead>
<tr>
<th>Test</th>
<th>df</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recessive against rest</td>
<td>11</td>
<td>P=0.02</td>
</tr>
<tr>
<td>Least dominant against dominant and most dominant</td>
<td>11</td>
<td>P=0.06</td>
</tr>
<tr>
<td>Dominant against most dominant</td>
<td>11</td>
<td>P=0.01</td>
</tr>
<tr>
<td>Combined test</td>
<td>33</td>
<td>P=0.0006</td>
</tr>
</tbody>
</table>
Table 4. The number of populations where a haplotype is present as a function of dominance class for data and simulations for two different population sizes (average over 10 simulation runs)

<table>
<thead>
<tr>
<th>Dominance level</th>
<th>Population presence observed</th>
<th>Population presence simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pop size 100</td>
</tr>
<tr>
<td>Recessive</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Least dominant</td>
<td>9.33</td>
<td>10.0</td>
</tr>
<tr>
<td>Dominant</td>
<td>9.33</td>
<td>9.5</td>
</tr>
<tr>
<td>Most dominant</td>
<td>7.73</td>
<td>6.9</td>
</tr>
</tbody>
</table>
Table 5. The observed haplotype frequencies of each dominance class averaged over populations and compared to expectations. These are deterministically calculated frequencies assuming independent populations and frequencies obtained in simulations (see text). For simulations, standard deviations over 15 independent runs are shown. For comparison the last column shows the deterministic frequencies in a panmictic population.

<table>
<thead>
<tr>
<th>Dominance class</th>
<th>Number of haplotypes</th>
<th>Observed frequencies</th>
<th>Deterministic (independent populations)</th>
<th>Simulations Pop size 100</th>
<th>Simulations Pop size 200</th>
<th>Deterministic (panmictic population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recessive</td>
<td>1</td>
<td>0.357</td>
<td>0.357</td>
<td>0.357±0.035</td>
<td>0.368±0.032</td>
<td>0.314</td>
</tr>
<tr>
<td>Least dominant</td>
<td>3</td>
<td>0.066</td>
<td>0.069</td>
<td>0.073±0.011</td>
<td>0.072±0.009</td>
<td>0.070</td>
</tr>
<tr>
<td>Dominant</td>
<td>3</td>
<td>0.047</td>
<td>0.053</td>
<td>0.047±0.009</td>
<td>0.051±0.010</td>
<td>0.053</td>
</tr>
<tr>
<td>Most dominant</td>
<td>11</td>
<td>0.028</td>
<td>0.025</td>
<td>0.025±0.003</td>
<td>0.024±0.003</td>
<td>0.029</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Map of Iceland with location of sampled populations. Elevation above sea level is indicated by colors showing 200m intervals. Major Glaciers are shown in speckled white. Modified version of map at http://en.wikipedia.org/wiki/Image:Map_of_Iceland.svg, Max Naylor del.

Figure 2: Phylogenetic tree (Minimum evolution, HKY substitution model) of the observed haplotypes plus two outgroups (Aly3 and Aly7). The different dominance classes of the haplotypes are marked by colors: Yellow: Recessive, green: least dominant, red: dominant, blue: most dominant. Bootstrap values > 85% are indicated

Figure 3: Observed haplotype frequencies in the 12 populations grouped by dominance level. Yellow: Recessive, green: least dominant, red: dominant, blue: most dominant. Population number is marked in upper right corner of each graph

Figure 4. Expected and observed total frequencies of haplotypes at each dominance level in each population. Expected frequencies were calculated from the deterministic equations assuming that the observed number of haplotypes is that in the given population

Figure 5. Fst values based on S-haplotypes between pairs of population samples as a function of their distance (dotted line: mean Fst)

Figure 6. A.: Observed haplotype frequencies in data pooled over the 12 populations. B. Mate availabilities experienced by each of the haplotypes averaged over populations, (weighted by the number of the given haplotype in each population)
Figure 7: Average mate availabilities based on observed genotype frequencies for the 12 population samples compared to the mate availabilities calculated assuming the haplotypes observed in the populations.
Figure 1:
Figure 2:

- **Least dominant**
  - AISRK06
  - AISRK14
  - AISRK18

- **Dominant**
  - AISRK31
  - AISRK12
  - AISRK16
  - AISRK25
  - AISRK37
  - AISRK38

- **Recessive**
  - AISRK01

- Aly7
- Aly3

Distance Scale: 0.05
Figure 3
Figure 4

Expected frequencies

Observed frequencies

- Recessive
- Least dominant
- Dominant
- Most dominant
Figure 5

![Figure 5](image-url)
Figure 6

**Total Frequency**

**Average MA**

Haplotype
Figure 7