Molecular Ecology Resources (2008) **8**, 295–298 doi: 10.1111/j.1471-8286.2007.01985.x

COMPUTER PROGRAMS

NESSI: a program for numerical estimations in sporophytic self-incompatibility genetic systems

S. BILLIARD
GEPV — Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Université des Sciences et Technologies de Lille 1, F-59655 Villeneuve d'Ascq cedex, France

Abstract

NESSI is a computer program generating predictions about allelic and genotypic frequencies at the S-locus in sporophytic self-incompatibility systems under finite and infinite populations. For any pattern of dominance relationships among self-incompatibility alleles, NESSI computes deterministic equilibrium frequencies and estimates distributions in samples from finite populations of the number of alleles at equilibrium, allelic and genotypic frequencies at equilibrium and allelic and genotypic frequency changes in a single generation. These predictions can be used to rigorously test the impact of negative frequency-dependent selection on diversity patterns in natural populations.

Keywords: allelic richness, frequency-dependent selection, genetic diversity, mating system

Received 2 July 2007; revision accepted 8 August 2007

Molecular ecologists are increasingly interested in testing the action of natural selection in natural populations. To perform this task, they need tools enabling them to disentangle the relative importance of various evolutionary pressures including genetic drift, and they need to compare their observations to theoretical predictions derived from the precise selective pressure acting on the gene(s) they investigate. Self-incompatibility in plants is a widespread genetic system found in nearly half of all plant families (de Nettancourt 2001) preventing inbreeding and especially self-fertilization in hermaphrodites through recognition and rejection of incompatible pollen, including self pollen. In two of the main plant families where the molecular bases of self-incompatibility have been described in detail, two tightly linked genes at the so-called S-locus encode interacting proteins determining pollen and pistil specificities. Two genotypes expressing the same specificity cannot cross. It is well known that this mating system results in negative frequency-dependent selection (FDS) on the S-locus, since an individual expressing a rare specificity has a higher number of potential mates. In species with gametophytic self-incompatibility (GSI), pollen specificity depends on its haploid genome and pistil specificity depends on its diploid genome with codominant interactions among alleles. In GSI, theoretical investigations have shown that if FDS is the only selective process acting on the S-locus, then all S-alleles should be equifrequent (Wright 1939) and allelic diversity should reflect an equilibrium between the rate of appearance of new specificities in the population and their loss by drift (Nagylaki 1975). Such predictions are more difficult to obtain in species with sporophytic self-incompatibility (SSI), first because the specificity expressed in pollen depends on the diploid genotype of the paternal plant, and second, dominance relationships can occur between each pair of alleles, both in pollen and pistil. Theoretical investigations have shown that the expected frequencies at equilibrium as well as the expected allelic diversity mainly depend on those dominance relationships (Bateman 1952; Schierup *et al*. 1997; Billiard *et al*. 2007). However, a simple way for computing precise expectations in SSI does not exist so far, except for much contrasted cases, not covering the full range of dominance patterns observed in natural populations (see Billiard *et al*. 2007 for a review). Yet, the number and frequency of S-alleles in species with SSI together with their pattern of dominance relationships are increasingly being investigated, for instance in *Sinapis arvensis* (Stevens & Kay 1989), *Ipomoea trifida* (Kowyama *et al*. 1994), *Arabidopsis lyrata* (Mable *et al*. 2003, Schierup *et al*. 2006) and *Brassica insularis* (Glémin *et al*. 2005).

We have implemented a program that performs such predictions based on the general model of SSI developed by Billiard *et al*. (2007) and computes (i) the deterministic equilibrium frequency of alleles in an infinite population;

Correspondence: Sylvain Billiard, Fax: +33 3 20 43 69 79; E-mail: sylvain.billiard@univ-lille1.fr

Table 1 Dominance relationships between alleles, observed allelic frequencies in an Icelandic population (Schierup *et al*. 2006) and expected frequency at deterministic equilibrium. Note that dominance relationships were considered identical in pistil and pollen but can be changed independently by the user

Allele name	S1	S2	S3	S4	S5	S6	T1	T2	U1	U2	V	Observed frequency	Deterministic expected frequency
S1	1	0.5	0.5	0.5	0.5	0.5	1	1	1	1	1	0.067	0.045
S2	0.5	1	0.5	0.5	0.5	0.5	1	1	1	1	1	0.09	0.045
S3	0.5	0.5	1	0.5	0.5	0.5	1	1	1	1	1	0.018	0.045
S4	0.5	0.5	0.5	1	0.5	0.5	1	1	1	1	1	0.012	0.045
S5	0.5	0.5	0.5	0.5	1	0.5	1	1	1	1	1	0.012	0.045
S6	0.5	0.5	0.5	0.5	0.5	1	1	1	1	1	1	0.006	0.045
T1	0	0	0	0	0	0	1	0.5	1	1	1	0.006	0.077
T2	0	0	0	0	0	0	0.5	1	1	1	1	0.108	0.077
U1	0	0	0	0	0	0	0	0	1	0.5	1	0.03	0.102
U2	0	0	0	0	0	0	0	0	0.5	1	1	0.275	0.102
V	0	0	0	0	0	0	0	0	0	0	1	0.377	0.371

(ii) the distribution of the number of alleles maintained in a finite isolated population as well as the distribution of allelic and genotypic frequencies; (iii) the distribution of allelic and genotypic frequency changes in a single generation in a finite isolated population. These predictions can be used to determine whether the observed frequencies in samples from natural populations are compatible with the hypothesis that FDS is the most important evolutionary process acting on the S-locus and therefore shapes the diversity patterns in natural populations.

Models and hypotheses

The population is composed of N diploid individuals, $N \rightarrow \infty$ for deterministic equilibrium computations. The population is assumed to be isolated and panmictic: pollination is free between individuals. Compatibility is determined by the program through the construction of a compatibility table (p_{ijkl}) with $p_{ijkl} = 1$ if a pollen produced by a genotype *ij* is compatible with a pistil produced by a genotype *kl*, $p_{ijkl} = 0$ otherwise.

The program is based on recurrent genotypic frequency changes across nonoverlapping generations: all plants grow, produce an infinite number of pollen grains and ovules, get pollinated, produce seeds and die. The program computes the frequency f_{ij} of genotype *ij* after reproduction (Billiard *et al*. 2007). The frequency f_{ij} is therefore the genotypic frequency among seeds. Equilibrium is defined as a minimal allelic frequency change between generations. Stochastic simulations are performed by sampling N diploid individuals among all seeds produced. By default, the initial genotypic frequencies for stochastic simulations are considered to reflect deterministic equilibrium. Finally, the mutation rate is defined per chromosome per generation, and assumed to follow the K-alleles model: the total number of possible allelic states is finite and a mutation will change an allele into one of the K-1 other potential allelic states with equal probability.

Computations

Deterministic equilibrium

This option allows the computation of expected genotypic and allelic frequencies at deterministic equilibrium with FDS only, for any number of alleles and pattern of dominance relationships (Table 1). NESSI can thus help determine whether the observed empirical frequencies fit theoretical expectations.

Distributions of allelic richness, allelic and genotypic frequencies in finite populations

These options can be used to predict diversity patterns at the S-locus in a finite population. It is especially useful as a mean to estimate confidence intervals around deterministic values for expected allelic frequencies in a sample when drift is allowed, and to test whether the observed frequencies are significantly different from the expectation.

Expected changes in genotypic and allelic frequencies in one generation in finite populations

Because FDS will tend to push alleles towards their deterministic equilibrium frequency, an allele whose frequency lies below this threshold should increase in frequency to the next generation, while an allele whose frequency lies above this threshold should decrease in frequency. The magnitude of this change is specific to each allele and depends mainly on dominance (Billiard *et al*. 2007). In order to test quantitatively whether allele frequency changes observed between two successive generations are compatible with these predictions, NESSI computes the distribution of the frequency changes in one generation given observed initial genotypic frequencies at time *t*. One can then determine if a given allele follows the prediction under FDS and drift only.

Table 2 The first results ever of NESSI: (A) Distribution of allelic frequencies across 1000 replicate simulations (10 000 generations each) in a 200-individual population with mutation rate per chromosome per generation set to 10^{-5}. The 95% confidence interval of the expected allelic frequencies at drift–mutation–FDS equilibrium is shown on the right. (B) Distribution of allelic richness for each dominance class

(A) Distribution of allelic frequencies																Frequency confidence intervals 95%
Allele name	0	[0, 0.05]	[0.05, 0.1]	[0.1, 0.15]	[0.15, 0.2]	[0.2, 0.25]	[0.25, 0.3]	[0.3, 0.35]	[0.35, 0.4]	[0.4, 0.45]	[0.45, 0.5]	[0.5, 0.55]	[0.55, 0.6]	[0.6, 0.65]	[0.65, 1]	
S1	607	62	240	89	2	0	0	0	0	0	0	0	0	0	0	0–0.15
S2	634	49	248	67	2	0	0	0	0	0	0	0	0	0	0	0–0.15
S3	610	50	244	91	5	0	0	0	0	0	0	0	0	0	0	0–0.15
S4	635	56	239	69	1	0	0	0	0	0	0	0	0	0	0	0–0.15
S5	581	48	264	101	6	0	0	0	0	0	0	0	0	0	0	0–0.15
S6	608	65	251	74	2	0	0	0	0	0	0	0	0	0	0	0–0.15
T1	289	44	278	306	77	6	0	0	0	0	0	0	0	0	0	0–0.2
T2	260	34	281	320	91	14	0	0	0	0	0	0	0	0	0	0–0.2
U1	335	15	113	242	180	101	13	1	0	0	0	0	0	0	0	0–0.25
U2	298	19	108	257	197	104	12	1	2	0	0	1	1	0	0	0–0.25
V	2	0	0	0	0	1	3	30	105	310	292	208	42	7	0	0.3–0.6

(B) Distribution of allele number per dominance class							
Allele number:	0	1	2	3	4	5	6
Class 1 -	0	80	546	345	27	2	0
Class 2 -	5	539	456				
Class 3 -	4	625	371				
Class 4 -	2	998					

Ecological and genetic parameters

The program is versatile enough to allow the implementation of a whole range of contrasted ecological and genetic parameters. For all computations, the user has to choose between two reproduction regimes: FDS occurs either through pollen only (the so-called Wright's model, Wright 1939) or through both pollen and pistil (the so-called fecundity selection model, Vekemans *et al*. 1998). For stochastic simulations, the user can specify population size, sample size, mutation rate, as well as the total number of possible allelic states. The dominance relationships are given for every pair of alleles in a square matrix in which element at row *i* and column *j* is set to 1 if allele *i* is dominant over allele *j*, 0 if it is recessive or 0.5 if both alleles are codominant (see examples in Billiard *et al*. 2007 and Table 1 for *Arabidopsis lyrata* from Schierup *et al*. 2006). Two such matrices must be given: one for pollen and one for pistil. In cases where dominance classes can be defined, dominance relationships can also be given to take this structure into account, in which case the user simply provides the number of alleles per dominance class. For instance, in *A. lyrata* (Table 1), the dominance relationships can be given as follows: (6, 2, 2, 1), which means that there are four dominance classes, the most dominant one with six alleles, the second and third most dominant with two alleles and the most recessive with a single allele.

Features

NESSI is written in C++ language. The compiled executable is available for Windows only and the source files can be downloaded at the following address: www.univ-lille1.fr/gepv/downloads/Softwares/Nessi/Nessi.htm. The executable and input files should be placed in a single folder, in which all output files will also be placed.

Example

We re-analysed the data from a natural *A. lyrata* population from Iceland, which exhibits sporophytic self-incompatibility (Schierup *et al*. 2006). The dominance relationships between alleles were identical in pistil and pollen (Table 1). Table 1 also shows the observed allelic frequencies in a sample of 87 individuals. Census population size was estimated based on the observation that very few individuals were found around the sampling area (Schierup *et al*. 2006). We can thus test whether the diversity pattern in the sampled population is compatible with a small population size (e.g. $N = 200$). We used NESSI with the fecundity selection model, to predict allelic frequencies at deterministic equilibrium (Table 1), their confidence intervals for $N = 200$, as well as the distribution of allelic richness per dominance class in a population with mutation rate set to 10^{-5} (Table 2). Distributions were computed using 1000 independent replicates of 10 000 generations. A sample of

87 individuals was drawn randomly among the 200 individuals at the last generation of each replicate. Table 2A shows the number of replicates among 1000 replicates, where the frequency of a given allele was observed in a range (x, y). The first column '0' corresponds to lost alleles. Table 1 shows slight differences between the observed frequency and their equilibrium deterministic predictions, except for the most recessive 'V' allele which shows a narrow match. NESSI enabled us to test whether the observed slight discrepancies were compatible with the expected variance due to drift and/or sampling or whether some other factors need to be invoked. As shown in Table 2, all allele frequencies were within their predicted confidence intervals. However, Table 2B also showed that six alleles were never maintained in the most dominant class (alleles from S1 to S6) in any of the 1000 simulation replicates. In the majority of replicates (546/1000), this class contained only two alleles. Thus, in most stochastic simulations, the predicted allelic richness in this class was lower than that observed in the real data. We concluded from these results that $N = 200$ is too small a population size to allow such a high number of alleles to be maintained in an isolated population with such dominance relationship patterns. We hypothesize that the observed diversity patterns may be due to a large amount of immigrating pollen and hence a much larger effective size of the population. Indeed, as shown by Schierup *et al.* (2000), small levels of migration among populations can severely impact the diversity and allelic richness within populations, at a locus under balancing selection.

Acknowledgements

I wish to thank V. Castric, V. Llaurens and X. Vekemans for program testing and comments on the manuscript, and L. Amsellem for informatics help. This work was supported by an ATIP-plus research grant from the CNRS.

References

Bateman AJ (1952) Self-incompatibility systems in angiosperms. 1. Theory. *Heredity*, **6**, 285–310.

Billiard S, Castric V, Vekemans X (2007) A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. *Genetics*, **175**, 1351–1369.

Glémin S, Gaude T, Guillemin ML *et al*. (2005) Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. *Genetics*, **171**, 279–289.

Kowyama Y, Takahasi H, Muraoka K *et al*. (1994) Number, frequency and dominance relationships of S-alleles in diploid *Ipomoea trifida*. *Heredity*, **73**, 275–283.

Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity*, **90**, 422–431.

Nagylaki T (1975) The deterministic behaviour of self-incompatibility alleles. *Genetics*, **79**, 545–550.

de Nettancourt D (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants*. Springer-Verlag, Berlin, Germany.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics*, **147**, 835–846.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research*, **76**, 51–62.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB (2006) Selection at work in self-incompatible *Arabidopsis lyrata*: mating patterns in a natural population. *Genetics*, **172**, 477–484.

Stevens JP, Kay QON (1989) The number, dominance relationships and frequencies of self-incompatibility alleles in a natural population of *Sinapis arvensis* L. in South Wales. *Heredity*, **62**, 199–205.

Vekemans X, Schierup MH, Christiansen FB (1998) Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution*, **52**, 19–29.

Wright S (1939) The distribution of self-sterility alleles in populations. *Genetics*, **24**, 538–552.