

Sex inheritance in gynodioecious species: a polygenic view

Bodil K. Ehlers^{1,*}, Sandrine Maurice² and Thomas Bataillon^{3,4}

¹Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, Building 540, 8000 Aarhus C, Denmark

²ISEM, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France

³UMR 1097 “Diversité et Génomes des Plantes Cultivées”, Domaine de Melgueil, 34130 Mauguio, France

⁴BiRC, University of Aarhus, Høegh-Guldbergs Gade, Building 090, 8000 Aarhus C, Denmark

Gynodioecy is defined as the coexistence of two different sexual morphs in a population: females and hermaphrodites. This breeding system is found among many different families of angiosperms and is usually under nucleo-cytoplasmic inheritance, with maternally inherited genes causing male sterility and nuclear factors restoring male fertility. Numerous theoretical models have investigated the conditions for the stable coexistence of females and hermaphrodites. To date, all models rest on the assumption that restoration of a given male sterile genotype is controlled by a single Mendelian factor. Here, we review data bearing on the genetic determinism of sex inheritance in three gynodioecious plant species. We suggest that restoration of male fertility is probably best viewed as a quantitative trait controlled by many loci. We develop a threshold model that accommodates an underlying polygenic trait, which is resolved at the phenotypic level in discrete sexual morphs. We use this model to reanalyse data in *Thymus vulgaris*, *Silene vulgaris* and *Plantago coronopus*. A simple Mendelian inheritance of sex determinism is unlikely in all three species. We discuss how our model can shed additional light on the genetics of restoration and point towards future efforts in the modelling of gynodioecy.

Keywords: cytoplasmic male sterility; quantitative threshold model; sex determination

1. INTRODUCTION

Plants with gynodioecious breeding systems are composed of populations where hermaphrodites, bearing perfect flowers, coexist with females. Gynodioecy is a widespread condition found among many angiosperm families. The sex determination mechanism of females and hermaphrodites is crucial for understanding the evolution and maintenance of this sexual polymorphism. For most gynodioecious species, a combination of cytoplasmic genes causing male sterility (hereafter CMS) and nuclear restorer alleles determine the sex of an individual plant (reviewed in Charlesworth 1981). In short, female plants carry a CMS gene blocking the development of functional anthers, while hermaphrodites either do not have such a CMS gene or carry at least one nuclear restorer allele that restores male function. Different CMS genes can coexist within a species and even within a population (e.g. Van Damme & van Delden 1982; Koelewijn & van Damme 1995a; Manicacci *et al.* 1996), each with its own specific nuclear restorer allele(s). The molecular mechanism responsible for male sterility is still unknown, but CMS has been attributed to a (chimeric) open reading frame (ORF) in the mitochondrial genome (e.g. in *Zea mays*, *Brassica napus*, *Petunia* spp.). CMS associated genes are thought to be derived from portions of known genes fused with unknown sequences (Wise & Pring 2002). These new ORFs encode proteins that are believed to interfere with

pollen production, causing male sterility. Nuclear restorers alter the expression of these CMS associated genes by modifying their transcripts and suppressing their deleterious effects on male function, thereby restoring male fertility (Schnable & Wise 1998; Bentolila *et al.* 2002; Wise & Pring 2002).

The important feature of such a sex determination system is that CMS genes are located in the mitochondria, and thus are usually maternally inherited (Reboud & Zeyl 1994). Therefore, a CMS gene can invade a population of hermaphrodites provided that females who carry it have even a slight female fitness advantage compared to the female fitness of hermaphrodites (Lewis 1941). Higher seed set in females than hermaphrodites is indeed often observed (e.g. review in Delph 1999). For the females to be maintained in the population, however, polymorphism at nuclear restorer loci is needed, as the fixation of a restorer allele would cause the female phenotype to disappear. Evidence of ‘hidden’ CMS genes, masked by the fixation of nuclear restorers is found in many crop species where CMS is recovered after crosses between different lines or species for example, maize, rice (*Oryza sativa*) and sunflower (*Helianthus annuus*) (Schnable & Wise 1998). Maintenance of a polymorphism for nuclear restorer alleles is, however, possible if one assumes a cost of carrying such an allele in an alien cytoplasm (e.g. Charlesworth 1981; Frank 1989; Gouyon *et al.* 1991). A nuclear restorer allele is then selected for when in combination with the cytoplasm it restores, and selected

* Author for correspondence (bodil.ehlers@biology.au.dk).

against otherwise. This generates frequency-dependent selection on both the CMS and nuclear restorer alleles. Empirical evidence for costs of restoration is still sparse but the existence of costs have been demonstrated in *Plantago lanceolata* (de Haan *et al.* 1997) and *Lobelia siphilitica* (Bailey 2002).

Most theoretical studies on the maintenance of nucleo-cytoplasmic gynodioecy have made the simplifying assumption that, despite variation in CMS genes, each CMS only needs one restorer allele in order to have its male function restored, and in its simplest form the nuclear restorer allele is assumed to be dominant to the maintainer allele (Charlesworth 1981; Gouyon *et al.* 1991; see Bailey *et al.* 2003 for the study of other patterns of dominance and Saur Jacobs & Wade 2003 for a recent review of the theory). However, empirical studies show that some CMS genes may need a combination of two or more nuclear restorer alleles in order to have their male function restored. For example, in maize the T-cytoplasm causing male sterility is restored by the joint action of the nuclear restorer allele *Rf2* in combination with one to three additional restorers (Schnable & Wise 1998; Wise *et al.* 1999). Koelewijn & van Damme (1995b) suggested that at least five restorer loci were involved in restoring male sterile cytoplasm of *Plantago coronopus*, and Dudle *et al.* (2001) found that sex determination in *L. siphilitica* seemed to vary among cytoplasm with one CMS gene being restored by a single dominant allele, while restoration of other CMS only could be explained by the action of several nuclear loci and/or epistatic effects.

Here we propose that several loci with additive effects may be involved in restoring a single CMS. The idea emerged when reanalysing data of experimental crosses of the gynodioecious *Thymus vulgaris* from Assouad (1972). Sex determination in *T. vulgaris* is cyto-nuclear. In *T. vulgaris* females, flowers vary in their anther development. Three different flower types, so-called 'b-', 'c-' and 'd'-females, can be distinguished depending on the development of the sterile anthers (Thompson *et al.* 2002). A female plant carries only one type of flower. Such flower polymorphisms have also been reported from other gynodioecious species and have been proposed to be simply due to the action of different CMS genes blocking the development of anthers at different stages (e.g. Van Damme & van Delden 1982). Thus female offspring of a female plant should all have the same flower type as their maternal plant, as they all carry the same cytoplasm. However, analysis of female offspring from the three types of in *T. vulgaris* shows that although the maternal flower type is the commonest phenotype in the female offspring, a female can produce female offspring with all three flower types. Moreover, the sex ratio in the offspring also varies among females of different flower types (figure 1). When pollinated by a single pollen donor, b-females produce more hermaphrodites than c-females (Fischer's exact test, $p < 0.0001$) which in turn produce more hermaphrodites than d-females (Fischer's exact test, $p < 0.0001$; see figure 1). The same pattern is also found in open pollinated b-, c- and d-females (data not shown). These results show that (i) females of *T. vulgaris* vary in their restoration ability and (ii) that the female types which morphologically resemble hermaphrodites most (type b) also have a higher proportion of hermaphrodites in their offspring compared to females whose flowers have

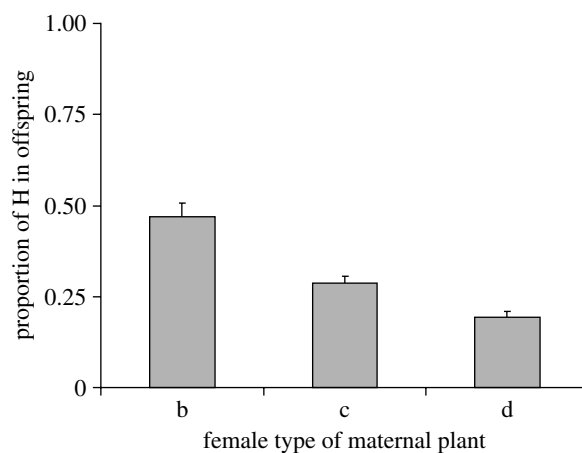


Figure 1. Mean proportion of hermaphrodites (\pm s.e.) in offspring of b-, c- and d-females (see text for further explanation) of *Thymus vulgaris*. Proportions are based on pooled offspring from 10 b-females ($n=344$ offspring), 17 c-females ($n=764$ offspring) and 18 d-females ($n=704$ offspring). Several pollen donors were used for these crosses, but each female was only crossed to a single pollen donor. Data extracted from Assouad (1972).

only little or no male organ vestiges. Gigord *et al.* (1999) also found that females with bigger corollas, which morphologically resemble hermaphrodites most, have a higher percentage of hermaphrodites in their progeny. The morphological gradation from d- (females with no male organs) to b-females (hermaphrodite-like) suggests that the restoration could be viewed as a quantitative trait controlled by several nuclear loci.

Below, we develop a threshold model for sex determination in gynodioecious species. We reanalyse data sets consisting of controlled crossing experiments from three different gynodioecious species (*P. coronopus*, *T. vulgaris* and *Silene vulgaris*). We show that our model can satisfactorily explain sex inheritance in such crosses. Finally, we discuss the value of our model and how it differs from Mendelian approaches adopted so far.

2. A THRESHOLD MODEL FOR SEX INHERITANCE

Threshold models have been widely used to model traits that are discrete at the phenotypic level but thought to be governed by polygenic inheritance. Traits whose inheritance have been successfully modelled through that framework include disease susceptibility, winglessness in insects (see Roff 1996 for a review) or more recently different sexual morphs (euphallic versus aphillic individuals) in snails (Ostrowski *et al.* 2000). Here, we assume that the discrete phenotypes (hermaphrodites and females) are controlled by an underlying trait, which we refer to as 'maleness'. Note that this trait is not actually observable. We assume that maleness is controlled by a large number of loci with additive gene action and is normally distributed in a population. The fraction of individuals that have a trait value higher than a given threshold, T , will have a hermaphrodite phenotype, whereas the remaining individuals will have a female phenotype (see figure 2a). Below, we outline our model for the case of a binary phenotype (female versus hermaphrodite), but the model can easily be extended to cases involving more than two phenotypes (see appendix).

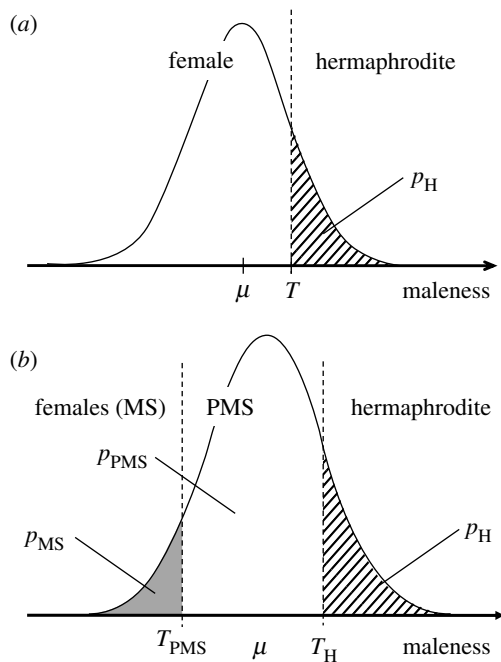


Figure 2. Graphical representation of the threshold model. (a) A two-morph threshold model with females and hermaphrodites. A single threshold T (dotted line) is used. (b). A three-morph threshold model featuring females, partially male sterile (PMS) and hermaphrodites. Two thresholds, T_{PMS} and T_H , are used.

(a) Predicted proportions of hermaphrodites in a cross

We consider a cross between (i) a female (or a hermaphrodite maternal parent) carrying the CMS type i having genetic value M_F^i and (ii) a (hermaphrodite) male parent with genetic value M_H^i . These two parents are randomly sampled from a population where maleness is normally distributed. Given the assumptions above, an offspring from that cross will inherit a genetic value, G , which is normally distributed with mean $\mu = 1/2(M_H^i + M_F^i)$ and variance σ^2 . The expected proportion of hermaphrodite phenotypes in such a cross is then

$$p_H = \int_T^{+\infty} \phi_G(x) dx, \tag{2.1}$$

where ϕ_G is the probability density function of a Normal distribution with mean μ and variance σ^2 .

(b) Maximum likelihood estimation of the parameters of the threshold model

A dataset (see below) will typically consists of a series of N crosses involving n male parents with genetic values $M_{H1}^i, M_{H2}^i, \dots, M_{Hn}^i$ crossed to a series of m female parents with values $M_{F1}^i, M_{F2}^i, \dots, M_{Fm}^i$. Let us 'label' the crosses c_1, c_2, \dots, c_N . Within the j th cross we observe nf_j females and nh_j hermaphrodites. If we assume that the threshold model outlined above governs inheritance of sexual phenotypes in such a cross, we can write the likelihood of such data as

$$l(c_j) = \binom{nf_j + nh_j}{nh_j} p_H^{nh_j} (1 - p_H)^{nf_j}, \tag{2.2}$$

where p_H is the probability defined by equation (2.1). The likelihood of the whole dataset is then (assuming N independent crosses)

$$l_D = \prod_{j=1}^N l(c_j).$$

Maximum likelihood estimates (hereafter MLE) of the parameters of the model can be obtained by maximizing l_D over the space of parameters. The parameters are the genetic values of the different parents involved in the crosses, i.e. M_F 's and M_H 's values, the within-family variance σ^2 and the value of the threshold T . Our focus here is predicting the sex ratio of progeny using our threshold model and not to estimate components of variance or the heritability of the maleness trait. We will restrict our analysis to crosses involving unrelated individual resulting in progeny that are assumed to be outbred, and thus assume that within-family variances are identical for a given species dataset. Without loss of generality we can set the within-family variance to $\sigma^2=1$ and set the threshold to $T=0$. Note that assuming a different value for T would only shift all genetic values accordingly but would not change the predicted sex ratios under the threshold model. If only a single cross was available as data, the genetic value of the mother, M_{F1} and the father M_{H1} of the cross would not be estimable, only their sum would (as it determines the mean of the distribution of within-family values for maleness). But as data from more independent crosses involving the same parents are available, all parental genetic values can be estimated (§3 for details on the type of crosses analysed for each species) provided that at least k crosses are available (if k parental genetic values are to be estimated). Note, however, that some maternal effects may be embedded into the maternal genetic value we estimate.

The likelihood of the data under the threshold model can be compared with a saturated model fitting the data perfectly. The saturated model is obtained by setting in equation (2.2) the p s to their observed value for each cross. Our threshold model is nested within this saturated model, and a likelihood ratio test, also known as a G -test, can therefore be used to assess the fit of our model relative to the saturated model. A large value of G indicates that the saturated model fits the data significantly better and that some of the variation in sex ratio is not accounted for by the threshold model. The G -statistic is distributed as a chi-square with d degrees of freedom, where d is the number of parameters fitted in the saturated model minus the number of parameters used by the threshold model. All calculations were carried out using MATHEMATICA 4.1 (Wolfram 1996) and likelihoods were maximized numerically using the function FindMinimum.

3. DATA ANALYSIS

(a) *Plantago coronopus*

The data of Koelewijn & van Damme (1995b) consist of controlled crosses made between plants sampled at two locations, 'KwadeHoek' (hereafter KH) and Oostvoornse meer (hereafter OM). Within each location, two ecologically different sites were surveyed: (i) KH, a dune a salt marsh and (ii) OM a dune and a meadow. At each site, three hermaphrodites and two females were sampled within an area of about 50 m². The following crosses were made (omitting data from selfing hermaphrodites that are not meeting the assumption of outbred progeny of our model).

Each female was crossed with every hermaphrodite used as a male parent resulting in 6 sets of offspring. Hermaphrodites were crossed with each other (including reciprocal crosses) resulting in 6 sets of offspring.

In total, there were thus 12 sets of crosses per site involving five different parents. The data used here are extracted from the subtables 'first generation crosses' of tables 1–4 from Koelewijn & van Damme (1995b). The maternal plants have two possible cytoplasms, which are, respectively, denoted by cytoplasm 1 and 2. Within each set of offspring, three categories of individual were recorded: (i) male sterile (MS or females), (ii) partially male sterile (PMS) and (iii) hermaphrodites (H). The categories PMS and H were pooled in a single category (as done by the authors in the original paper).

The threshold model was fitted by estimating eight parameters (three genetic values of the hermaphrodites for each of the two cytoplasms and the two genetic values of the mothers). These genetic values were then used to predict sex ratios within each cross. Our threshold model predicts well the proportion of hermaphrodites for all but a few crosses from the last sampling site (OM meadow; table 1). Accordingly, the saturated model does not fit the data better than our reduced model in two of the four sites (table 1).

(b) *Thymus vulgaris*

We reanalysed data from the study of Belhassen *et al.* (1991), also analysed by Charlesworth & Laporte (1998) using several Mendelian models. The data consisted of crosses of nine female plants all descended from the same maternal plant, i.e. all carrying the same cytoplasm, crossed to six hermaphrodites from the same population as the female plants. Not all possible crosses among the 15 plants (nine females, six hermaphrodites) were performed but, each hermaphrodite was crossed with four of the nine females (see table 5 in Belhassen *et al.* 1991) and the sex ratio of the offspring was recorded. Using available data from all 24 independent crosses, we fitted our model by estimating the genetic value of female and hermaphrodite plants (one for each parent so 15 parameters in total). This allowed us to obtain the sex ratios predicted under the threshold model for comparison with the observed ones. The model predicts the sex ratio very well, except for crosses with very few offspring (table 2; cross c10 and c17). Over all, due to those crosses where predictions from the threshold model are off by a few percent, the saturated model fits the data better than our reduced model.

(c) *Silene vulgaris*

Charlesworth & Laporte (1998) analysed data from several crossing experiments involving individuals originating from two different populations of *S. vulgaris*. The data we could use to apply to the threshold model from that study was, however, limited as individuals had to be involved in several different crosses in order to be able to estimate their parental genetic value. Given these requirements, five crosses (crosses 18–22 from table 6 in Charlesworth & Laporte 1998) were used. Our threshold model was fitted using four parameters (one for each parent involved in the crosses). The threshold model provides a good fit between observed and expected proportion of hermaphrodites in all five crosses, and the saturated model does not fit the data significantly better (table 3).

Table 1. Observed (obs. *H*) and expected (exp. *H*) values of proportion of hermaphrodites in offspring of *Plantago coronopus* under the threshold model.

(Data was obtained from tables 1–4 in Koelewijn & van Damme (1995b); *n*, number of offspring in each cross; cross no. refers to the numbering used in the original presentation of the data. $G_{\text{threshold}}$ is the statistic comparing the fit of the threshold model relative to a saturated model fitting the data perfectly. Within each site, $G_{\text{threshold}}$ is distributed as a chi-square with 4 degrees of freedom (12 crosses minus eight parameters fitted in the threshold model).)

cross	exp. <i>H</i>	obs. <i>H</i>	<i>n</i>	$G_{\text{threshold}}$
<i>site: KH-dune</i>				
c107	0.999	0.977	43	4.401
c109	1	1	75	0
c108	0.978	1	75	3.337
c111	1	1	62	0
c110	1	1	79	0
c112	0.932	0.932	59	0
c35	1	1	70	0
c36	0.96	0.986	70	1.589
c37	0.472	0.437	71	0.358
c38	1	1	72	0
c39	0.802	0.806	72	0.006
c40	0.809	0.809	68	0
total				9.710 ($p=0.046$)
<i>site: KH-salt march</i>				
c101	0.998	1	66	0.264
c103	1	1	55	0
c102	1	1	59	0
c105	0.887	0.9	10	0.0175
c104	1	1	61	0
c106	0.769	0.768	181	0.001
c29	0.567	0.565	62	0.002
c30	0.38	0.382	68	0.002
c31	0.781	0.781	73	0
c32	0.833	0.833	72	0
c33	1	1	72	0
c34	1	1	70	0
total				0.287 ($p>0.5$)
<i>site: OM-dune</i>				
c83	1	1	76	0
c85	0.897	0.865	133	1.385
c84	1	1	66	0
c87	1	1	78	0
c86	0.665	0.7335	75	1.634
c88	1	1	64	0
c23	0.908	0.971	68	4.237
c24	0.45	0.45	60	0
c25	0.688	0.612	67	1.737
c26	0.944	0.944	72	0.0003
c27	0.185	0.185	54	0
c28	0.531	0.531	64	0
total				8.99 ($p=0.06$)
<i>site: OM-meadow</i>				
c77	0.469	0.469	81	0
c79	0.976	1	69	3.352
c78	1	1	78	0
c81	0.967	1	62	4.161
c80	0.999	0.986	72	3.418
c82	0.999	0.987	76	3.317
c17	0.221	0.027	75	24.06
c18	0.973	1	71	3.89
c19	0.852	1	76	24.35
c20	0.682	0.826	69	7.367
c21	0.999	1	75	0.150
c22	0.989	0.901	71	18.83
total				92.9*** ($p<0.001$)

Table 2. Observed (obs.) and expected (exp.) values of proportion of hermaphrodites in offspring of *Thymus vulgaris* under the threshold model.

(Data was obtained from Belhassen *et al.* (1991). Cross no. refers to the number of family numbers used Charlesworth & Laporte (1998); n , number of offspring in each cross. $G_{\text{threshold}}$ is the statistic comparing the fit of the threshold model relative to a saturated model fitting the data perfectly. $G_{\text{threshold}}$ is distributed as a chi-square with 9 degrees of freedom (24 crosses minus 15 parameters fitted in the threshold model).)

cross	exp. H	obs. H	n	$G_{\text{threshold}}$
c1	0.0235	0	44	2.093
c2	0.029	0.091	22	1.935
c3	0.021	0	9	0.382
c4	0.0346	0.031	32	0.011
c5	0.269	0.379	29	1.671
c6	0.333	0.268	41	0.802
c7	0.461	0.379	29	0.789
c8	0.362	0.412	34	0.358
c9	0.148	0.143	21	0.004
c10	0.195	0.286	7	0.332
c11	0.299	0.292	24	0.006
c12	0.218	0.2	20	0.039
c13	0.174	0.222	9	0.136
c14	0.031	0	41	2.582
c15	0.161	0.182	11	0.034
c16	0.257	0.290	31	0.176
c17	0.077	0.5	2	2.516
c18	1	0.922	51	7.678
c19	0.0698	0.072	57	0.0001
c20	0.127	0.097	31	0.275
c21	0.232	0	13	6.863
c22	0.059	0.091	44	0.699
c23	0.414	0.468	47	0.561
c24	0.319	0.289	38	0.155
total				30.10 ($p < 0.001$)

4. DISCUSSION

(a) *Patterns of sex inheritance in Plantago, Thymus and Silene*

In all three species, the models suggested by the authors as best fitting the observed sex ratio, all depart from a single Mendelian factor. A model with two loci was used to fit the *T. vulgaris* data (Charlesworth & Laporte 1998), three loci in the case of *S. vulgaris* (Charlesworth & Laporte 1998) and up to five loci were needed in the case of *P. coronopus* (Koelewijn & van Damme 1995b). All authors presented goodness of fit test (G -tests) to assess how well their Mendelian model of sex inheritance fitted the actual data. Accordingly, we also assessed the fit of our threshold model relative to a saturated model.

However, there is no simple way of comparing the fit of the threshold model with the Mendelian models used previously, because the two classes of models are not nested. Akaike's information criteria (Burnham & Anderson 1998) could potentially be used to compare these. But, such comparisons do not make sense here as the previous analysis performed by the authors actually proceeded by sequentially rejecting a number of simple Mendelian models, adding loci and/or dominance and epistasis until a satisfactory fit was achieved. It is then very hard to decide whether the last model retained fits the data robustly or if the achieved fit was partly due to chance.

Table 3. Observed (obs.) and expected (exp.) values of proportion of hermaphrodites in offspring of *Silene vulgaris* under the threshold model.

(Data obtained from Charlesworth & Laporte (1998; table 6). Cross no. refers to the family number used in Charlesworth & Laporte (1998); n , sample size in number of offspring. $G_{\text{threshold}}$ is the statistic comparing the fit of our threshold model relative to a saturated model fitting the data perfectly. $G_{\text{threshold}}$ is distributed as a chi-square with 1 degree of freedom (five crosses minus four parameters fitted in the threshold model).)

cross	exp. H	obs. H	n	$G_{\text{threshold}}$
c18	0.08	0.08	25	0
c19	0.799	0.75	24	0.340
c20	0.67	0.72	25	0.291
c21	0.08	0.12	25	0.477
c22	0.036	0	25	1.818
total				2.93 ($p = 0.09$)

The procedure is best described by Koelewijn & Van Damme (1995b, p. 1771): 'Because the minimum overall model gave a bad fit, we have to assume the involvement of further (population specific) genes. All possible crosses can then be explained, but for every deviant cross a new locus has to be assumed. The number of genes involved thus increases rapidly, approaching polygenic determination of sex'. Instead of comparing the relative fits of the two classes of models we concentrate on the 'absolute' fit achieved by both types of approach.

Overall, assuming that restoration of cytoplasmic male sterility is a quantitative trait, does at least as good a job in explaining the data, in terms of absolute goodness of fit, as assuming that restoration is a classical Mendelian trait (compare observed and predicted proportions of hermaphrodites under the threshold model and G -tests in tables 1–3). In the *Plantago* data, the threshold model predicts accurately the observed sex ratios in three out of the four sites. In the last site (OM-dune), sex ratios predicted by the threshold model are off by 15–20% compared to the observed sex ratios in three of the 12 crosses where the Mendelian fits the data much better. This may be due to the fact that highly skewed sex ratios, very close to either 0 or 1, are not adequately captured by the Gaussian assumption of the threshold model when assuming homogeneous within-family variance. In the *Thymus* data, ignoring crosses with very few offspring, observed sex ratios are accurately predicted. Finally, the threshold model fits the sex ratios in the *Silene* dataset well whereas Mendelian models performed rather poorly in two of the five crosses reanalysed here (see table 6 in Charlesworth & Laporte 1998).

(b) *Biological insights on the restoration of CMS*

The data from *P. coronopus* consists of offspring from crosses involving two different cytoplasm. This gives us the opportunity to examine whether the restoration of cytoplasm 1 and 2 can be viewed as two different traits. We built two nested models that make different assumptions regarding restoration of male fertility. Model 1 assumes that the genetic value of a father changes with the cytoplasm into which his pollen is applied—this may be a reasonable assumption if the nuclear genes have different restoration effects depending on the cytoplasm it is crossed

Table 4. Likelihood of *Plantago* datasets under two competing threshold models for sex inheritance.

(l_1 , likelihood of data under threshold model 1 which was fitted using eight free parameters (two genetic values for each of the three hermaphrodite and two genetic values for the females). l_2 , likelihood of data under threshold model 2 which was fitted with five free parameters (one genetic value for each parent). G : log likelihood ratio test for comparisons of model 1 and 2. G is asymptotically distributed as a chi-square with three degrees of freedom.)

dataset	l_2	l_1	G
KH-dune	1.974×10^{-8}	8.993×10^{-14}	24.6 ($p < 0.0001$)
KH-meadow	3.810×10^{-6}	1.657×10^{-11}	24.7 ($p < 0.0001$)
OM-dune	9.108×10^{-10}	1.839×10^{-11}	7.8 ($p = 0.05$)
OM-meadow	1.19×10^{-28}	4.389×10^{-42}	61 ($p < 0.0001$)

with. Model 2 is nested within model 1 and assumes that the genetic value of a father is the same no matter which cytoplasm the pollen is crossed with. Table 4 summarizes the likelihood of the *P. coronopus* data under the two competing models. The G -statistic reveals that in three sites out of four, the model 1 fits the data far better than the alternative. In the fourth sampling site, the reduced model that assumes a single genetic value for each father fits the data extremely well, leaving little room for improvement using model 1.

The estimated genetic values of hermaphrodites on cytoplasm 1 show no correlation with those of the same hermaphrodites on cytoplasm 2 ($r = 0.041$, $p > 0.1$). If a positive correlation was found, hermaphrodites good at restoring cytoplasm 1 should also be good at restoring cytoplasm 2 and thus the restoration of the two cytoplasm could be regarded as a single trait. The absence of a correlation between the genetic values, suggest that restoration of the two cytoplasm may be viewed as two independent traits. Note that sampling errors around the estimates of genetic values may reduce the power to detect correlations between the two traits. However, comparison of the relative fit of model 1 versus model 2 (table 4; LRT) indicates that the full model is often much better at explaining the data. This supports the idea that the ability to restore cytoplasm 1 is indeed independent of the ability to restore a different type of cytoplasm 2.

Our threshold model can also be used to estimate genetic values and thus characterize individuals with respect to their restoration ability. Such information is important in designing future crosses both in crop improvement where some breeding programmes routinely use cytoplasmic male sterilities and to generate testable predictions to investigate further patterns of sex inheritance in gynodioecious species.

(c) Extending the threshold model to incorporate intermediate morphs

In many gynodioecious species intermediate forms are reported. In *T. vulgaris* the b-type female could represent such an intermediate form. In *Beta vulgaris* two types of females coexist, the female type with more developed stamens has, like in *Thymus*, more hermaphrodites in its progeny (Boutin-Stadler *et al.* 1989). In *Plantago* species partial male steriles (PMS) are often reported in significant numbers (e.g. van Damme & van Delden 1982, Koelewijn & van Damme 1995b), and in *S. vulgaris* both PMS and female plants with different levels of anther development have been found (Charlesworth & Laporte 1998). The significance of these intermediate morphs for the evolution of cytoplasmic male sterility has not yet been

investigated in detail, and they are usually just pooled with the female or the hermaphrodite morph in the analysis of crosses. Our threshold model provides a natural framework to take into account such morphs. The extension of the threshold model from two morphs to three morphs is straightforward (see appendix). Briefly, we still use a single underlying trait (termed maleness) but use two thresholds. For a maleness measure lower than the first threshold, the male sterile phenotype will be produced. For maleness between this and the next threshold, PMS will be observed and hermaphrodites correspond to individuals with maleness above the second threshold (see figure 2b). We used Koelewijn & van Damme's data on *P. coronopus* for the number on intermediates and analysed the data under the three morph threshold model. The model predicts well the proportion of females, hermaphrodites and intermediates (see appendix).

In conclusion, we have developed a threshold model to study sex inheritance in gynodioecious species. Use of that model to analyse patterns of inheritance in three different species show that viewing restoration of cytoplasmic male sterility as a quantitative trait is an alternative approach to understanding the often complicated restoration patterns of cytoplasmic male sterility. Moreover, the threshold model also allows us to deal with the PMS which are largely neglected when fitting Mendelian inheritance models.

Viewing restoration of male sterility as a polygenic trait calls for further research examining how different genetic determinism of sex inheritance shapes selection on CMS and restorer alleles. Models of the maintenance of gynodioecy have assumed so far that there are large fitness differences between sexual morphs and strong frequency dependent selection on nuclear restorer alleles. Alternative models assuming that restoration is a gradual process where fitness gains and costs are spread over a relatively large number of loci throughout the genome could potentially lead to different dynamics dominated by weak selection.

We thank M. H. Schierup, F. B. Christiansen and D. Charlesworth for their constructive criticism and comments on the previous version of this manuscript.

APPENDIX. EXTENSION OF THE THRESHOLD MODEL TO THREE MORPHS

Three ordered categories of sexual phenotypes are assumed: male sterile, partially male sterile and hermaphrodites. These categories correspond to the data of Koelewijn & van Damme (1995b).

Let p_H , p_{PMS} , p_{MS} be the probability of producing, respectively, a hermaphrodite, a partially male sterile

phenotype and a male sterile phenotype. That probability is a function of the mean value of the family for maleness as well as two thresholds T_{PMS} and T_{H} . Assuming that maleness is normally distributed with mean μ and variance σ^2 , the expected proportion of each morph are

$$\left. \begin{aligned} p_{\text{H}} &= \int_{T_{\text{H}}}^{+\infty} \phi_G(x) dx, \\ p_{\text{PMS}} &= \int_{T_{\text{PMS}}}^{T_{\text{H}}} \phi_G(x) dx, \\ p_{\text{MS}} &= \int_{-\infty}^{T_{\text{PMS}}} \phi_G(x) dx. \end{aligned} \right\} \quad (\text{A } 1)$$

Here ϕ_G is the density of a normal distribution with mean μ and variance σ^2 .

If within the j th cross n_1 male sterile, n_2 partially male sterile and n_3 hermaphrodites are observed, the likelihood associated with that dataset is

$$\ell(c_j) = \frac{(n_1 + n_2 + n_3)!}{n_1! n_2! n_3!} p_{\text{MS}}^{n_1} p_{\text{PMS}}^{n_2} p_{\text{H}}^{n_3}, \quad (\text{A } 2)$$

where the p s are the probabilities defined by equation (A 1). The likelihood of the whole dataset is then (assuming N independent crosses)

$$\ell_{\text{D}} = \prod_{j=1}^N \ell(c_j). \quad (\text{A } 3)$$

As for the two morphs model, MLE of the parameters of the model can be obtained by maximizing ℓ_{D} over the space of relevant parameters (the parental genetic values and the thresholds).

cross		MS	PMS	H	n
c107	E	0.002	0.03	0.97	43
	O	0.02	0.21	0.77	
c109	E	0.001	0.02	0.98	75
	O	0	0.08	0.92	
c108	E	0.04	0.17	0.79	75
	O	0	0.15	0.85	
c111	E	0.04	0.17	0.79	62
	O	0	0.08	0.92	
c110	E	0.04	0.16	0.80	79
	O	0.0	0.14	0.86	
c112	E	0.06	0.21	0.73	59
	O	0.07	0.24	0.69	
c35	E	0.05	0.19	0.76	70
	O	0	0.31	0.69	
c36	E	0.08	0.23	0.69	70
	O	0.01	0.09	0.90	
c37	E	0.36	0.36	0.28	71
	O	0.56	0.21	0.23	
c38	E	0.02	0.12	0.86	72
	O	0	0.07	0.93	
c39	E	0.12	0.29	0.59	72
	O	0.19	0.16	0.65	
c40	E	0.21	0.34	0.45	68
	O	0.19	0.50	0.31	

Observed (O) and expected (E) proportions of male sterile (MS), partially male sterile (PMS) and hermaphrodite (H) individuals in *P. coronopus* at the KH-dune site under a three morphs model.

REFERENCES

- Assouad, M. W. 1972 Recherche sur la génétique écologique de *Thymus vulgaris* L. Thèse de Doctorat es Sciences, USTL, Montpellier, France.
- Bailey, M. 2002 A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* **56**, 2178–2186.
- Bailey, M. F., Delph, L. F. & Lively, C. M. 2003 Modelling gynodioecy: novel scenarios for maintaining polymorphism. *Am. Nat.* **161**, 762–776.
- Belhassen, E., Dommée, B., Atlan, A., Gouyon, P. H., Pomente, D., Assouad, M. W. & Couvet, D. 1991 Complex determination of male sterility in *Thymus vulgaris* L.: genetic and molecular analysis. *Theor. Appl. Genet.* **82**, 137–143.
- Bentolila, S., Alfonso, A. A. & Hanson, M. R. 2002 A pentatricopeptide repeat containing gene restores fertility to cytoplasmic male sterile plants. *Proc. Natl Acad. Sci. USA* **99**, 10 887–10 892.
- Boutin-Stadler, V., Saumitou-Laprade, P., Valero, M., Jean, R. & Vernet, P. 1989 Spatio-temporal variation of male sterile frequencies on two natural populations of *Beta maritima*. *Heredity* **63**, 395–400.
- Burnham, K. P. & Anderson, D. R. 1998 *Model selection and inference—a practical information-theoretic approach*. New York: Springer.
- Charlesworth, D. 1981 A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* **46**, 27–39.
- Charlesworth, D. & Laporte, V. 1998 The male-sterility polymorphism of *Silene vulgaris*: analysis of genetic data from two populations and comparison with *Thymus vulgaris*. *Genetics* **150**, 1267–1282.
- de Haan, A. A., Hundscheid, M. P. J. & van Hinsberg, A. 1997 Effects of CMS types and restorer alleles on plant performance in *Plantago lanceolata* L.: an indication for cost of restoration. *J. Evol. Biol.* **10**, 803–820.
- Delph, L. 1999 Sexual dimorphism in life history. In *Gender and sexual dimorphisms in flowering plants* (ed. M. A. Geber, T. E. Dawson & L. F. Delph), pp. 149–173. Berlin: Springer.
- Dudle, D. A., Mutikainen, P. & Delph, L. F. 2001 Genetics of sex determination in the gynodioecious species *Lobelia siphilitica*: evidence from two populations. *Heredity* **86**, 265–276.
- Frank, S. A. 1989 The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* **133**, 345–376.
- Gigord, L., Lavigne, C. J., Shykoff, J. & Atlan, A. 1999 Evidence for the effects of restorer genes on male and female reproductive functions of hermaphrodites in the gynodioecious species *Thymus vulgaris* L. *J. Evol. Biol.* **12**, 596–604.
- Gouyon, P. H., Vichot, F. & Van Damme, J. M. M. 1991 Nuclear cytoplasmic male sterility: single point equilibria versus limit cycles. *Am. Nat.* **137**, 498–514.
- Koelewijn, H. P. & van Damme, J. M. M. 1995a Genetics of male sterility in gynodioecious *Plantago coronopus*. 1. Cytoplasmic variation. *Genetics* **139**, 1749–1758.
- Koelewijn, H. P. & van Damme, J. M. M. 1995b Genetics of male sterility in gynodioecious *Plantago coronopus*. 2. Nuclear genetic variation. *Genetics* **139**, 1759–1775.
- Lewis, D. 1941 Male sterility in natural populations of hermaphrodite plants. *New Phytol.* **40**, 56–63.
- Manicacci, D., Couvet, D., Belhassen, E., Gouyon, P. H. & Atlan, A. 1996 Founder effects and sex ratio in the gynodioecious *Thymus vulgaris* L. *Mol. Ecol.* **5**, 63–72.
- Ostrowski, M. F., Jarne, P. & David, P. 2000 Quantitative genetics of sexual plasticity: the environmental threshold

- model and genotype-by-environment interaction for phallus development in the snail *Bulinus truncatus*. *Evolution* **54**, 1614–1625.
- Reboud, X. & Zeyl, C. 1994 Organelle inheritance in plants. *Heredity* **72**, 132–140.
- Roff, D. A. 1996 The evolution of threshold traits in animals. *Q. Rev. Biol.* **71**, 3–35.
- Saur Jacobs, M. & Wade, M. J. 2003 A synthetic review of the theory of gynodioecy. *Am. Nat.* **161**, 837–851.
- Schnable, P. S. & Wise, R. P. 1998 The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci.* **3**, 175–180.
- Thompson, J. D., Rolland, A. G. & Prugnolle, F. 2002 Genetic variation for sexual dimorphism in flower size within and between populations of gynodioecious *Thymus vulgaris*. *J. Evol. Biol.* **15**, 362–372.
- Van Damme, J. M. M. & van Delden, W. 1982 Gynodioecy in *Plantago lanceolata* L. Polymorphism for plasmon type. *Heredity* **49**, 303–318.
- Wise, R. P. & Pring, D. R. 2002 Nuclear-mediated mitochondrial gene regulation and male fertility in higher plants: light at the end of the tunnel? *Proc. Natl Acad. Sci. USA* **99**, 10 240–10 242.
- Wise, R. P., Gobelman-Werner, K., Pei, D., Dill, C. L. & Schnable, P. S. 1999 Mitochondrial transcript processing and restoration of male fertility in T-cytoplasm maize. *J. Hered.* **90**, 380–385.
- Wolfram, S. 1996 *The mathematica book*, 3rd edn. Cambridge University Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.