Mating systems and selection efficacy: a test using chloroplastic sequence data in Angiosperms

S. GLÉMIN* & A. MUYLE†

*Institut des Sciences de l’Évolution de Montpellier, UMR CNRS 5554, Montpellier, France
†Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Lyon, France

Keywords:
chloroplast;
deleterious mutations;
dN/dS ratio;
mating systems;
selection;
selling.

Abstract

Selfing is assumed to reduce selection efficacy, especially purifying selection. This can be tested using molecular data, for example by comparing the Dn/Ds ratio between selfing and outcrossing lineages. So far, little evidence of relaxed selection against weakly deleterious mutations (as inferred by a higher Dn/Ds ratio) in sellers as compared to outcrossers has been found, contrary to the pattern often observed between asexual and sexual lineages. However, few groups have been studied to date. To further test this hypothesis, we compiled and analysed chloroplastic sequence data sets in several plant groups. We found a general trend towards relaxed selection in sellers in our data sets but with weak statistical support. Simulations suggested that the results were compatible with weak-to-moderate Dn/Ds ratio differences in selfing lineages. Simple theoretical predictions also showed that the ability to detect relaxed selection in sellers could strongly depend on the distribution of the effects of deleterious mutations on fitness. Our results are compatible with a recent origin of selfing lineages whereby deleterious mutations potentially have a strong impact on population extinction or with a more ancient origin but without a marked effect of deleterious mutations on the extinction dynamics.

Introduction

The evolution of selfing from outcrossing is one of the most frequent life-history trait transitions in angiosperms (Stebbins, 1957) and may also have recurrently occurred in many other groups of animals (Jarne & Auld, 2006) and fungi (Billiard et al., 2011; Gioti et al., 2012). Transitions from obligate outcrossing (self-incompatible species) to self-compatibility and eventually to selfing are mainly irreversible (Igic et al., 2006, 2008; Igic & Busch, 2013) and come with higher extinction rates (Goldberg et al., 2010). Selfing is thus supposed to be an evolutionary dead-end strategy (Stebbins, 1957; Takebayashi & Morrell, 2001; Igic & Busch, 2013). Although there is increasing evidence in favour of the dead-end hypothesis, the underlying causes of higher extinction rates in sellers are still being debated (Glémin & Galtier, 2012; Wright et al., 2013).

Selfing automatically reduces the effective population size, \( N_e \), by two-fold because of nonindependent gamete sampling during reproduction (Pollak, 1987; Nordborg, 1997), while also reducing effective recombination because it mainly occurs between homozygote sites (Nordborg, 2000). Genetic hitchhiking effects due to higher genetic linkage (Maynard-Smith & Haigh, 1974; Charlesworth et al., 1993) are thus expected to reduce \( N_e \) below the automatic two-fold level. Finally, recurrent bottlenecks that are considered to be more frequent in sellers since a single seed can find a new population (Schoen & Brown, 1991; Ingvarsson, 2002) can also further reduce \( N_e \). The overall reduction in \( N_e \) could be formulated as follows (Glémin, 2007):

\[
N_e (F) = \frac{\alpha(F)}{1 + F} \frac{N}{1 + F}
\]

where \( F \) is Wright’s fixation index \( (F_{IS}) \), and \( 0 < \alpha(F) \leq 1 \) summarizes the hitchhiking and bottleneck effects. The main hypotheses put forward to explain the dead-end theory are thus based on the premise that,

Correspondence: Sylvain Glémin, Institut des Sciences de l’Évolution.
CC64, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.
Tel.: +33 (0)4 67 14 35 87; fax: +33 (0)4 67 14 36 10;
e-mail: sylvain.glemin@univ-montp2.fr

© 2014 THE AUTHORS. J. EVOL. BIOL.
JOURNAL OF EVOLUTIONARY BIOLOGY © 2014 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY
because of the low $N_e$ and weak recombination efficiency, selfing should reduce the selection efficacy, potentially leading to the accumulation of mildly deleterious mutations (Lynch et al., 1995; Schultz & Lynch, 1997; Glémin, 2007) and limiting adaptation to new biotic and abiotic environments (Stebbins, 1957; Agrawal & Lively, 2001; Morran et al., 2011; Glémin & Ronfort, 2013). However, the respective roles of reduced adaptive ability versus genetic deterioration are still unclear. Importantly, selfing also exposes alleles in homozygotes, thus increasing their apparent dominance levels and facilitating both positive and negative selection. For intermediate dominance (i.e. additive selection, whereby the homozygotic effect of a mutation is twice the heterozygotic effect), the apparent dominance is increased by a $1 + F$ factor, which exactly compensates for the automatic $N_e$ reduction (Caballero & Hill, 1992; Charlesworth, 1992; Pollak & Sabran, 1992). If mutations are recessive or partially recessive, as presumed for deleterious mutations, $N_e$ must be reduced beyond the two-fold level – that is, $\omega(F) < 1$ – for selection to be reduced in selfers (Glémin, 2007). This is also true for the fixation of new beneficial mutations, but not for adaptation from standing variation, which is less dependent on dominance levels, and which is more efficient in outcrossers than in selfers (Glémin & Ronfort, 2013).

The accumulation of deleterious mutations can be tested through molecular approaches by comparing the rates of nonsynonymous versus synonymous change in divergence ($D_{N}/D_{S}$, hereafter) or polymorphism ($P_n/P_s$). If most changes in amino acids are neutral or deleterious, $\omega$ and $P_n/P_s$ should be higher in selfing lineages due to the relaxation of selection against these deleterious mutations. Such approaches have been conducted in several species (Table 1) and gave mixed results. Most studies based on divergence failed to detect relaxed selection in selfers, except in *Neurospora*, where tiny differences in $\omega$ ($\omega_{\text{out}} \sim 0.14$ vs. $\omega_{\text{in}} \sim 0.17$) have been detected using a very large data set of more than 2700 genes (Gioti et al., 2012). On the contrary, most studies based on polymorphism detected relaxed selection in selfers compared to outcrossers. This suggests that selfing, hence relaxed selection, is of recent origin and can only be properly detected on a short time scale. This has been clearly illustrated in a study of the recently derived selfing species *Capsella rubella*, where a new method to identify founding haplotypes of ancestral populations allowed separation of ancestral and novel variations (Brandvain et al., 2013). The authors detected a strong signature of relaxed selection in newly emerged variation (high $P_n/P_s$ ratio within founding haplotypes) compared to ancestral variation (low $P_n/P_s$ ratio between founding haplotypes).

In contrast, studies comparing asexual and sexual species more often detected relaxed selection in asexual lineages using divergence measures, even with relatively small data sets (Table 1). Note, however, that the assumption that higher $\omega$ values in asexual lineages can be directly attributed to clonality has been recently challenged in a new analysis of *Daphnia* genomes (Tucker et al., 2013). This suggests that deleterious mutations should accumulate at a slower rate in sellers than in asexuals and could not be the main cause of the extinction of the former (see review and discussion in Glémin & Galtier, 2012). However, data are still insufficient to draw a firm conclusion, especially because the increase in $\omega$ in selfing lineages has not been tested in many plant groups.

Here, we have therefore extended the testing of this hypothesis to new plant groups, while focusing on two chloroplastic genes (*matK* and *rbcL*) for several reasons. These genes have been widely sequenced for phylogenetic studies and are thus available for many angiosperm species, especially in groups whose mating systems have been mapped onto a phylogeny. As chloroplastic genomes are haploid, they are not subject to dominance effects, thus sidestepping possible confounding effects of homozygosity on selection efficacy (see above). The two-fold reduction in $N_e$ due to homozygosity does not affect haploid genomes so that only bottleneck and hitchhiking effects do matter:

$$N_e^{\text{chlor}}(F) = \pi(F)N_{\text{female}}$$

As stated above, this additional more than two-fold reduction in $N_e$ ($\pi(F) < 1$) is pivotal to the argument that selection is less efficient in selfers (Glémin, 2007). Bottlenecks should affect cytoplasmic and nuclear genes similarly, whereas nuclear genes should be more sensitive to hitchhiking effects. Nuclear genes are affected by selection due to sites physically linked on the same chromosome and also to sites on other chromosomes that are genetically – but not physically – linked through selfing. Conversely, selection on the nuclear genome affects chloroplastic genes only through linkage due to selfing. For species with few chromosomes and/or a short nuclear genetic map, $\pi(F)$ should be higher for chloroplastic than for nuclear genes. However, for species with numerous chromosomes and/or a long nuclear genetic map, physical linkage is negligible for nuclear genes and $\pi(F)$ should be similar for chloroplastic and for nuclear genes. This could be more formally illustrated by comparing $\pi(F)$ under the background selection model for chloroplastic and nuclear genes (see Appendix 1), as shown in Fig. 1. The difference is substantial only for very short genetic maps (dotted lines equivalent to only one chromosome of 100 cM). For larger maps, differences between nuclear and chloroplastic genes can occur when the genomic mutation rate is high and average selection coefficient against deleterious mutations is rather low (Fig. 1c). Another potential problem of our approach is that the data sets are limited to only two genes (but also see some studies.
Table 1 Summary of studies comparing molecular evolution patterns between selfing and outcrossing or asexual and sexual species (updated from Glémin & Galtier, 2012).

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Groups compared</th>
<th>Data set</th>
<th>dN/dS</th>
<th>pN/pS</th>
<th>Positive selection</th>
<th>Codon usage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selfing vs. outcrossing</td>
<td>Angiosperms 29 selfers/42 outcrossers</td>
<td>Meta-analysis (polymorphism)</td>
<td>±</td>
<td>±</td>
<td></td>
<td></td>
<td>Glémin et al. (2006)</td>
</tr>
<tr>
<td>Arabidopsis 1 selfer/1 outcrosser</td>
<td>23 nuc genes + 1 chloro gene</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>Wright et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>Arabidopsis 1 selfer/1 outcrosser</td>
<td>675/62 nuc genes</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>Foxe et al. (2008)</td>
<td></td>
</tr>
<tr>
<td>Arabidopsis/Brassica</td>
<td>1 selfer/2 outcrossers</td>
<td>185 nuc genes</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>Wright et al. (2007)</td>
</tr>
<tr>
<td>Arabidopsis/Capsella</td>
<td>1 selfer/1 outcrosser</td>
<td>257 nuc genes</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>Slotte et al. (2010)</td>
</tr>
<tr>
<td>Arabidopsis/ Capsella</td>
<td>2 selfers/2 outcrossers</td>
<td>780, 89, 120, 257 nuc genes</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Qiu et al. (2011)</td>
</tr>
<tr>
<td>Capsella</td>
<td>1 selfer/1 outcrosser</td>
<td>Complete genome/transcriptome</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Slotte et al. (2013)</td>
</tr>
<tr>
<td>Caenorhabditis</td>
<td>2 selfers/4 outcrossers</td>
<td>&gt; 1000 nuc genes</td>
<td>-</td>
<td>±</td>
<td></td>
<td></td>
<td>Cutter et al. (2008)</td>
</tr>
<tr>
<td>Collinsia</td>
<td>1 selfer/1 outcrosser</td>
<td>17 nuc genes + transcriptomes</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td>Hazzouri et al. (2013)</td>
</tr>
<tr>
<td>Eichhornia</td>
<td>3 selfers/1 outcrosser</td>
<td>~ 8000 nuc genes</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td>Ness et al. (2012)</td>
</tr>
<tr>
<td>Neurospora</td>
<td>32 homothallic/17 heterothallic</td>
<td>7 nuc genes</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Nygren et al. (2011)</td>
</tr>
<tr>
<td>Neurospora</td>
<td>1 pseudohomothallic/1 heterothallic</td>
<td>&gt; 2000 nuc genes</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>Whittle et al. (2011)</td>
</tr>
<tr>
<td>Neurospora</td>
<td>4 homothallic/31 heterothallic</td>
<td>&gt; 2700 nuc genes</td>
<td>+</td>
<td>±</td>
<td></td>
<td></td>
<td>Gioti et al. (2013)</td>
</tr>
<tr>
<td>Triticeae</td>
<td>2 selfers/2 outcrossers</td>
<td>52 nuc genes + 1 chloro gene</td>
<td>-</td>
<td>±</td>
<td></td>
<td></td>
<td>Haudry et al. (2008)</td>
</tr>
<tr>
<td>Triticeae</td>
<td>9 selfers/10 outcrossers</td>
<td>27 nuc genes</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>Escobar et al. (2010)</td>
</tr>
<tr>
<td>Asexuals vs. sexuals</td>
<td>Aphis</td>
<td>4 sexuals/4 asexuals</td>
<td>255 nuc genes + 10 mito genes</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campeloma</td>
<td>6 asexuals/12 sexuals</td>
<td>1 mito gene (Cytb)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Johnson &amp; Howard (2007)</td>
</tr>
<tr>
<td>Daphnia</td>
<td>14 asexuals/14 sexuals</td>
<td>Complete mito genome</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Paland &amp; Lynch (2006)</td>
</tr>
<tr>
<td>Daphnia</td>
<td>11 asexuals/11 sexuals</td>
<td>Complete mito genome</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>Tucker et al. (2013)</td>
</tr>
<tr>
<td>Oenothera</td>
<td>16 asexuals/16 sexuals</td>
<td>1 nuc gene (chI9)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Hersch-Green et al. (2012)</td>
</tr>
<tr>
<td>Potamopyrgus</td>
<td>14 asexuals/14 sexuals</td>
<td>Complete mito genome</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Neiman et al. (2010)</td>
</tr>
<tr>
<td>Rottlers</td>
<td>3 asexuals/2 sexuals</td>
<td>1 nuc gene (Hsp 82)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>Mark Welch &amp; Meselson (2001)</td>
</tr>
<tr>
<td>Rottlers</td>
<td>3 asexuals/4 sexuals</td>
<td>1 mito gene (Cox I)</td>
<td>±</td>
<td>+</td>
<td></td>
<td></td>
<td>Barracough et al. (2007)</td>
</tr>
<tr>
<td>Timema</td>
<td>6 asexuals/7 sexuals</td>
<td>2 nuc genes + 1 mito gene</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>Henry et al. (2012)</td>
</tr>
</tbody>
</table>

*Terminal vs. internal branches not controlled. Positive results are in bold.

in asexuals, Table 1), which limits the statistical power of individual analyses. However, any general trends should be detected by the combination of data sets. We thus also conducted simulations to test the power of our analyses to detect reductions in the selection efficacy in selfers.

Materials and methods

Data sets

We built the data sets by combining two sources of information. First, we searched in the literature for publications in which the mapping of contrasted mating systems on a phylogeny was documented. We used information from studies comparing selfers and outcrossers, self-compatible and self-incompatible species, and homostylous and heterostylous species. Hereafter, SELF denotes any one of selfing, self-compatible or homostyous mating system, and OUT refers to any outcrossing, self-incompatible or heterostyous mating system. Although direct tests of the hypothesis of the accumulation of deleterious mutations should concern selfing vs. outcrossing species, the self-compatibility status is usually associated with rather high selfing rates, and it could be a short intermediate step towards selfing (Igic et al., 2008; Igic & Busch, 2013). For instance, this rationale has been used to test the dead-end hypothesis in Solanaceae (Goldberg et al., 2010). We then conducted a GenBank search for matK and rbcl sequences corresponding to the species present in these phylogenies. Sequences were aligned with MUSCLE (Edgar, 2004) and manually checked and cleaned. When sequence lengths were too heterogeneous between species, we kept the species with the longest sequences to avoid numerous alignment gaps.

Previously published trees served as reference trees for each group of species. When sequences were lacking, we removed the corresponding species from
the corresponding tree. Methods implemented to map mating system evolution on phylogenies may differ between publications. To homogenize data sets and determine the status of all branches, as required in dN/dS analyses, we opted to map mating systems using parsimony and assuming unidirectional shifts from outcrossing (resp. self-incompatibility or heterostyly) to selfing (resp. self-compatibility or homostyly). The resulting maps were usually very close to those proposed in the original publications.

The list of species with the corresponding GenBank accessions is given in Table S1. Phylip files of alignments and trees with mapped mating systems are provided in Dryad repository.

Sequence analyses

We used the codeml program of the paml package (Yang, 2007) to perform various tests of codon evolution along phylogenies. Before testing the effect of mating systems on \( \omega \), we assessed the possible occurrence of positive selection in the sequences, as it has been shown that positive selection can affect rbcL (Kapralov & Filatov, 2007). We used the ‘site models’ implemented in codeml to detect potential sites under positive selection. We compared the M7 model (with a beta distribution of \( \omega \) values between 0 and 1) and the M8 model (with a beta distribution plus an additional category with \( \omega > 1 \)) by a likelihood ratio test (LRT) with two degrees of freedom. When sites under positive selection were detected, we built a second data set by removing from the alignment the sites belonging to the \( \omega > 1 \) class with high posterior probability (Bayes empirical Bayes prob. > 0.95).

For each data set, including those without sites under positive selection, we ran several nested ‘branch models’ (Fig. 2). We first ran a null model with a single \( \omega \) for all branches (\( M_{0\text{-null}} \)) and a model with two \( \omega \), that is, one for SELF branches and one for OUT branches (including internal branches; \( M_{\text{self-out}} \) Fig. 2a). However, changes in terminal branches may correspond to polymorphic mutations, not substitutions, yielding higher \( \omega \) because weakly deleterious mutations can be transiently polymorphic before eventually being lost. This can be misleading because selfing, SC or homostylous is mainly found on terminal branches. Moreover, assignation of mating systems on internal branches can be less accurate than on terminal branches. We thus ran two alternative models: a model with two \( \omega \), one for internal branches (\( M_{\text{int-ext}} \)) and a model with three \( \omega \), one for internal branches, one for SELF external branches and one for OUT external branches (Fig. 2c). Finally, some data sets presented internal SELF branches. We thus ran a model with four ratios for SELF internal, OUT internal, SELF external and OUT external branches (\( M_4 \) Fig. 2d). Nested models were tested by LRTs with the appropriate number of degrees of freedom.

To combine information from the different data sets, we performed binomial sign tests to compare \( \omega \)
between internal and external branches and between SELF and OUT branches. We also combined likelihood between data sets for each model by summing log-likelihoods. As we wanted to perform unilateral tests, that is, $\omega_{\text{ext}} > \omega_{\text{int}}$ and $\omega_{\text{SELF}} > \omega_{\text{OUT}}$, we added the log-likelihood of the alternative model of a given data set only if the order of $\omega$ values corresponded to the alternative hypothesis. Otherwise, if the order of $\omega$ was contrary to that expected, we added the log-likelihood of the corresponding null model. In such cases, the data set did not contribute to increasing the likelihood of the alternative model but cost one degree of freedom (Escobar et al., 2010).

Simulations

To assess the power of our analyses to detect relaxed selection in sellers, we performed simulations to determine which differences in $\omega$ were detectable in our analyses. For simplicity, we chose two reference trees with only 12 species: one with six SELF species on the leaves and one with a clade of three SELF species (Fig. 3). In both trees, the total length of SELF branches corresponded to 25% of the total tree. We only considered two $\omega$ and did not distinguish between internal and external branches. We used the evolver program of the paml package (Yang, 2007) to simulate sequences along these two phylogenies with two different $\omega$. We simulated sequences of 999 bp (333 codons) with a codon composition and a $\text{Ts/Tv}$ parameter (=2.37) corresponding to the matK data set in Veronica. To vary the simulated data set sizes, we set the total tree length at 0.5, 1 and 5. We used three $\omega$ values for the OUT branches: $\omega_{\text{OUT}} = 0.1, 0.25$ and 0.5. For the SELF branches, we increased the $\omega_{\text{OUT}}$ values by the following quantities $\Delta \omega = \omega_{\text{SELF}} - \omega_{\text{OUT}} = 0.05, 0.1, 0.15, 0.2$ and 0.25. We simulated 100 data sets for every combination of parameters (two trees, three total tree lengths, three $\omega_{\text{OUT}}$ and five $\omega_{\text{SELF}}$). Then, we ran the $M_0$ and $M_{\text{self-out}}$ models defined above with codeml.
We recorded the $\omega_{\text{SELF}}$ and $\omega_{\text{OUT}}$ values estimated in the $M_{\text{self-out}}$ model and the $P$-value of the LRT of the two models.

## Results

We obtained 19 data sets corresponding to 13 species groups representative of several angiosperm families (Table 2). The smallest data set contained 10 species with a 466 bp alignment and the largest contained 175 species with a 1515 bp alignment. Three data sets (for Exochaenium and Schiedea genera) had fewer than 50 substitutions (NdN + SdS in Table 2), so some groups of branches had very few or no synonymous and/or non-synonymous substitutions, and some models had an aberrant or undefined $\omega$ ratio. The results are presented in Table 3, but these three data sets were not taken into further account in the analyses.

The two genes mostly evolved under purifying selection. They both had $\omega < 1$, and, on average, $rbcL$ was more constrained than $\text{matK}$ (mean $\omega = 0.20$ vs. $\omega = 0.45$). External branches also showed higher $\omega$ than internal branches in 15 data sets of 16 (unilateral sign test: $P = 0.0003$; combined LRT: $\chi^2 = 60.34$, d.f. = 16, $P < 10^{-6}$, Table 3), which is the expected pattern under purifying selection. However, significant positive selection was found in most data sets (12 of 16) still showed higher external $\omega$ than $\omega_{\text{OUT}}$. Similar results were obtained when codons potentially evolving under positive selection were removed (second data sets), but the power to detect differences by LRT was lower (Table 4): $\omega_{\text{SELF}}$ was higher than $\omega_{\text{OUT}}$ for 13 of 16 data sets (unilateral sign test: $P = 0.010$; combined LRT $\chi^2 = 18.48$, d.f. = 16, $P = 0.297$), and the difference was significant for only two different data sets (Pontederiaceae $rbcL$ and Primula $rbcL$). However, as the SELF branches were mostly terminal, the previous results could be explained by the differences between internal and external branches that we detected in most data sets (see above). We thus tested the difference between mating systems on external branches only (models $M_3$ vs. $M_{\text{int-ext}}$). A majority of data sets (11 of 16) still showed higher external $\omega_{\text{SELF}}$ than $\omega_{\text{OUT}}$, but the difference was not significant (unilateral sign test: $P < 0.05$).

### Table 2 Summary of the data sets used in the analyses: taxonomic group, mating system type.

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>MS type</th>
<th>No. of S, SC, or homo</th>
<th>No. of O, Sl or het</th>
<th>Gene</th>
<th>No. of sites</th>
<th>NdN</th>
<th>SdS</th>
<th>No. of codons $\omega &gt; 1$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asteraceae</td>
<td>SC/Sl</td>
<td>55</td>
<td>18</td>
<td>matK</td>
<td>1536</td>
<td>442.2</td>
<td>266.2</td>
<td>2</td>
<td>Ferrer &amp; Good-Avila (2007)</td>
</tr>
<tr>
<td></td>
<td>SC/Sl</td>
<td>19</td>
<td>11</td>
<td>rbcL</td>
<td>1455</td>
<td>144.4</td>
<td>226.3</td>
<td>9</td>
<td>Fiz et al. (2008)</td>
</tr>
<tr>
<td>Exochaenium</td>
<td>homo/het</td>
<td>10</td>
<td>3</td>
<td>matK</td>
<td>466</td>
<td>31.5</td>
<td>18.5</td>
<td>0</td>
<td>Bena et al. (1996)</td>
</tr>
<tr>
<td>Geraniaceae</td>
<td>S/O</td>
<td>48</td>
<td>14</td>
<td>rbcL</td>
<td>1425</td>
<td>132.9</td>
<td>98.2</td>
<td>8</td>
<td>Barrett et al. (1996)</td>
</tr>
<tr>
<td>Linum</td>
<td>homo/het</td>
<td>11</td>
<td>6</td>
<td>rbcL</td>
<td>1425</td>
<td>16.2</td>
<td>106.6</td>
<td>0</td>
<td>Armbruster et al. (2006)</td>
</tr>
<tr>
<td>Medicago</td>
<td>S/O</td>
<td>31</td>
<td>25</td>
<td>matK</td>
<td>1518</td>
<td>225.9</td>
<td>108.5</td>
<td>6</td>
<td>Kohn et al. (1996)</td>
</tr>
<tr>
<td>Polemoniaceae</td>
<td>S/O</td>
<td>50</td>
<td>19</td>
<td>matK</td>
<td>1074</td>
<td>400.1</td>
<td>221.2</td>
<td>8</td>
<td>EMIN AND A. MUYLE</td>
</tr>
<tr>
<td>Pontederiaceae</td>
<td>homo/het</td>
<td>23</td>
<td>8</td>
<td>rbcL</td>
<td>1344</td>
<td>35.9</td>
<td>145.2</td>
<td>3</td>
<td>Mast et al. (2006)</td>
</tr>
<tr>
<td>Primula</td>
<td>homo/het</td>
<td>175</td>
<td>50</td>
<td>matK</td>
<td>1515</td>
<td>1127.8</td>
<td>702.2</td>
<td>10</td>
<td>Escobar et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>homo/het</td>
<td>45</td>
<td>9</td>
<td>rbcL</td>
<td>1416</td>
<td>196.7</td>
<td>323.2</td>
<td>8</td>
<td>Muller &amp; Albach (2010)</td>
</tr>
<tr>
<td>Psychotria</td>
<td>homo/het</td>
<td>15</td>
<td>4</td>
<td>matK</td>
<td>864</td>
<td>98.5</td>
<td>64.4</td>
<td>0</td>
<td>Sakai &amp; Wright (2008)</td>
</tr>
<tr>
<td>Schiedea</td>
<td>S/O</td>
<td>27</td>
<td>6</td>
<td>rbcL</td>
<td>553</td>
<td>40.8</td>
<td>30.8</td>
<td>4</td>
<td>Sakai et al. (2006)</td>
</tr>
<tr>
<td>Solanaceae</td>
<td>S/O</td>
<td>22</td>
<td>4</td>
<td>matK</td>
<td>1077</td>
<td>24.9</td>
<td>7.3</td>
<td>1</td>
<td>Escobar et al. (2010)</td>
</tr>
<tr>
<td>Triticeae</td>
<td>S/O</td>
<td>20</td>
<td>7</td>
<td>rbcL</td>
<td>1305</td>
<td>31.1</td>
<td>81.5</td>
<td>1</td>
<td>Muller &amp; Albach (2010)</td>
</tr>
</tbody>
</table>

SI, self-incompatible; SC, self-compatible; homo, homostyly; het, heterostyly; S, selfing; O, outcrossing; total, total number of species; NdN, total number of nonsynonymous substitutions; SdS, total number of synonymous substitutions.
Table 3 Results of $\omega = \text{Dn/Ds}$ analyses for branch models.

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Gene</th>
<th>$M_3$</th>
<th>$M_{3\text{est}}$ vs. $M_3$</th>
<th>$M_{3\text{self-out}}$ vs. $M_{3\text{self}}$</th>
<th>$P$-values</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Polemoniaceae</em></td>
<td>matK</td>
<td>0.441</td>
<td>0.430 0.451 0.455 0.445 0.528</td>
<td>$\chi^2 = 22.98, \text{d.f.} = 16$; only two data sets were individually significant (<em>Polemoniaceae</em> $rbcL$ and <em>Solanaceae</em> $rbcL$). Similar results are obtained for the second data sets: external $\omega_{\text{SELF}}$ was higher than external $\omega_{\text{OUT}}$ for 11 of 16 data sets (unilateral sign test: $P = 0.105$; combined LRT $\chi^2 = 11.5, \text{d.f.} = 16, P = 0.077$), and only one data set was individually significant (<em>Polemoniaceae</em> $rbcL$).</td>
<td>0.005 0.021 0.021</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>matK</td>
<td>0.446</td>
<td>0.425 0.519 0.635 0.528</td>
<td>$\chi^2 = 4204.60, \text{d.f.} = 16$</td>
<td>0.273 0.312</td>
</tr>
<tr>
<td><em>Verbascum</em></td>
<td>matK</td>
<td>0.285</td>
<td>0.293 0.485</td>
<td><em>2082.63</em> 0.422 0.515</td>
<td>0.001 0.001</td>
</tr>
<tr>
<td><em>Veronica</em></td>
<td>matK</td>
<td>0.361</td>
<td>0.337 0.544</td>
<td>$\chi^2 = 5316.93, \text{d.f.} = 16$</td>
<td>0.109 0.098</td>
</tr>
<tr>
<td><em>Linum</em></td>
<td>matK</td>
<td>0.303</td>
<td>0.206 0.481</td>
<td>$\chi^2 = 5139.24, \text{d.f.} = 16$</td>
<td>0.045 0.058</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>rbcL</td>
<td>0.389</td>
<td>0.285 0.565</td>
<td>$\chi^2 = 1435.17, \text{d.f.} = 16$</td>
<td>0.040 0.049</td>
</tr>
<tr>
<td><em>Polemoniaceae</em></td>
<td>rbcL</td>
<td>0.336</td>
<td>0.225 0.565</td>
<td>$\chi^2 = 3074.20, \text{d.f.} = 16$</td>
<td>0.339 0.472</td>
</tr>
<tr>
<td><em>Primula</em></td>
<td>matK</td>
<td>0.172</td>
<td>0.121 0.211</td>
<td>$\chi^2 = 4825.85, \text{d.f.} = 16$</td>
<td>0.005 0.021</td>
</tr>
<tr>
<td><em>Primula</em></td>
<td>rbcL</td>
<td>0.172</td>
<td>0.121 0.211</td>
<td>$\chi^2 = 4825.85, \text{d.f.} = 16$</td>
<td>0.005 0.021</td>
</tr>
<tr>
<td><em>Primula</em></td>
<td>matK</td>
<td>0.436</td>
<td>0.398 0.496</td>
<td>$\chi^2 = 2013.39, \text{d.f.} = 16$</td>
<td>0.539 0.755</td>
</tr>
<tr>
<td><em>Psychotria</em></td>
<td>matK</td>
<td>0.395</td>
<td>0.211 0.667</td>
<td>$\chi^2 = 1192.78, \text{d.f.} = 16$</td>
<td>0.022 0.196</td>
</tr>
<tr>
<td><em>Schiedea</em></td>
<td>matK</td>
<td>0.442</td>
<td>0.393 0.477</td>
<td>$\chi^2 = 5139.24, \text{d.f.} = 16$</td>
<td>0.054 0.058</td>
</tr>
<tr>
<td><em>Linum</em></td>
<td>rbcL</td>
<td>0.316</td>
<td>0.238 0.338</td>
<td>$\chi^2 = 1437.27, \text{d.f.} = 16$</td>
<td>0.482 0.025</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>rbcL</td>
<td>0.316</td>
<td>0.238 0.338</td>
<td>$\chi^2 = 1437.27, \text{d.f.} = 16$</td>
<td>0.482 0.025</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>rbcL</td>
<td>0.395</td>
<td>0.211 0.667</td>
<td>$\chi^2 = 1192.78, \text{d.f.} = 16$</td>
<td>0.022 0.196</td>
</tr>
<tr>
<td><em>Psychotria</em></td>
<td>matK</td>
<td>0.436</td>
<td>0.398 0.496</td>
<td>$\chi^2 = 2013.39, \text{d.f.} = 16$</td>
<td>0.539 0.755</td>
</tr>
<tr>
<td><em>Primula</em></td>
<td>matK</td>
<td>0.442</td>
<td>0.393 0.477</td>
<td>$\chi^2 = 5139.24, \text{d.f.} = 16$</td>
<td>0.054 0.058</td>
</tr>
<tr>
<td><em>Linum</em></td>
<td>rbcL</td>
<td>0.316</td>
<td>0.238 0.338</td>
<td>$\chi^2 = 1437.27, \text{d.f.} = 16$</td>
<td>0.482 0.025</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>rbcL</td>
<td>0.316</td>
<td>0.238 0.338</td>
<td>$\chi^2 = 1437.27, \text{d.f.} = 16$</td>
<td>0.482 0.025</td>
</tr>
<tr>
<td><em>Solanaceae</em></td>
<td>rbcL</td>
<td>0.395</td>
<td>0.211 0.667</td>
<td>$\chi^2 = 1192.78, \text{d.f.} = 16$</td>
<td>0.022 0.196</td>
</tr>
<tr>
<td><em>Psychotria</em></td>
<td>matK</td>
<td>0.436</td>
<td>0.398 0.496</td>
<td>$\chi^2 = 2013.39, \text{d.f.} = 16$</td>
<td>0.539 0.755</td>
</tr>
<tr>
<td><em>Primula</em></td>
<td>matK</td>
<td>0.411</td>
<td>0.348 0.533</td>
<td>$\chi^2 = 1621.56, \text{d.f.} = 16$</td>
<td>0.011 0.074</td>
</tr>
<tr>
<td><em>Schiedea</em></td>
<td>rbcL</td>
<td>0.622</td>
<td>0.406 0.584</td>
<td>*$\chi^2 = 2072.63, \text{d.f.} = 16$</td>
<td>0.001 0.001</td>
</tr>
</tbody>
</table>

M$_3$ one $\omega$; M$_{3\text{est}}$, internal and external branches, two $\omega$; M$_{3\text{self-out}}$, SELF and OUT branches, two $\omega$; M$_{3\text{self}}$, internal branches; SELF and OUT external branches, three $\omega$; M$_{4}$, SELF and OUT internal, SELF and OUT external branches, four $\omega$; InL, log-likelihood. The combined likelihood and $P$-value computations are explained in the main text. $P$-value lower than 0.05 are in bold.

*Undefined $\omega$ ratio.
Finally, we found neither individual nor general effects of mating systems on internal branches (Tables 3 and 4). Overall, we noted a general trend towards a higher $\omega$ in SELF branches, but the statistical support was weak.

Simulations

The simulation results were very similar between the topologies tested (Fig. 3), so we only present the results for the first topology with only SELF terminal branches (Fig. 4). The alternative topology was slightly more favourable for the detection of relaxed selection in SELF lineages (not shown). As expected, the power to detect differences between SELF and OUT branches increased with the total tree length. Throughout the different simulations, the average number of substitutions was 59 for the total tree length $= 0.5$, 118 for the total tree length $= 1$ and 595 for the total tree length $= 5$. Most of the real data sets were thus encompassed between the grey and black dots in Fig. 4 (see Table 2 for substitution values). Note, however, that this power was reduced for highly constrained genes with low $\omega$ ratio.

Table 4 As in Table 3 for data sets without sites under significant positive selection.

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Gene</th>
<th>$M_3$</th>
<th>$M_{int-ext}$</th>
<th>$M_{self-out}$</th>
<th>$P$-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\omega$</td>
<td>$\omega_{int-out}$</td>
<td>$\omega_{self-out}$</td>
<td>$\omega_{int-ext}$</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>matK</td>
<td>0.313</td>
<td>0.439</td>
<td>0.525</td>
<td>$-5463.88$</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>rbcL</td>
<td>0.098</td>
<td>0.139</td>
<td>0.167</td>
<td>$-3618.39$</td>
</tr>
<tr>
<td>Geraniaceae</td>
<td>rbcL</td>
<td>0.037</td>
<td>0.077</td>
<td>0.075</td>
<td>$-4950.60$</td>
</tr>
<tr>
<td>Medicago</td>
<td>matK</td>
<td>0.233</td>
<td>0.268</td>
<td>0.423</td>
<td>$-3963.72$</td>
</tr>
<tr>
<td>Polemoniaceae</td>
<td>matK</td>
<td>0.442</td>
<td>0.390</td>
<td>0.474</td>
<td>$-4808.06$</td>
</tr>
<tr>
<td>Pontederiaceae</td>
<td>rbcL</td>
<td>0.053</td>
<td>0.013</td>
<td>0.099</td>
<td>$-2652.73$</td>
</tr>
<tr>
<td>Primula</td>
<td>matK</td>
<td>0.388</td>
<td>0.422</td>
<td>0.403</td>
<td>$-1313.91$</td>
</tr>
<tr>
<td>Primula</td>
<td>rbcL</td>
<td>0.051</td>
<td>0.154</td>
<td>0.171</td>
<td>$-4311.96$</td>
</tr>
<tr>
<td>Psychotria</td>
<td>rbcL</td>
<td>0.120</td>
<td>0.305</td>
<td>0.462</td>
<td>$-1016.88$</td>
</tr>
<tr>
<td>Solanaceae</td>
<td>matK</td>
<td>0.348</td>
<td>0.326</td>
<td>0.480</td>
<td>$-4738.85$</td>
</tr>
<tr>
<td>Solanaceae</td>
<td>rbcL</td>
<td>0.000</td>
<td>0.180</td>
<td>0.331</td>
<td>$-1280.42$</td>
</tr>
<tr>
<td>Triticaceae</td>
<td>matK</td>
<td>0.096</td>
<td>0.395</td>
<td>0.254</td>
<td>$-2944.98$</td>
</tr>
<tr>
<td>Veronica</td>
<td>rbcL</td>
<td>0.050</td>
<td>0.248</td>
<td>0.129</td>
<td>$-2390.55$</td>
</tr>
<tr>
<td>Combined†</td>
<td></td>
<td>$-62865.46$</td>
<td>$-34756.60$</td>
<td>$-34756.60$</td>
<td>$-34756.60$</td>
</tr>
</tbody>
</table>

†For species without codons under positive selection, values of Table 3 were used to compute the combined likelihoods. *Undefined $\omega$ ratio.
To what variation in \(N_e\) – that is, to what \(a\) in eqn (1) – does a given \(r_o\) ratio correspond to? Assuming the distribution of fitness effects of deleterious mutations (DFEM) follows a gamma distribution with mean \(c = Ns\) and shape \(b\), we can show that (see Appendix 2):

\[
r_o = \frac{\omega_{\text{SELF}}}{\omega_{\text{OUT}}} \approx a^{-\beta}
\]  

For instance, for \(\beta = 0.25\), \(r_o = 1.5\) corresponds to \(a = 0.2\) if selfing occurred at the same time as speciation. However, if we assume that selfing only appeared on the last \(f\) fraction of the branches, we simply have:

\[
r_o \approx 1 - f + f a^{-\beta}
\]  

For instance, for \(\beta = 0.25\) and \(a = 0.2\) as above, \(r_o\) drops to 1.1 if selfing actually corresponded to only 20\% of the end of SELF branches, as could be the case in *Arabidopsis thaliana* (Tang et al., 2007; but see Bechsgaard et al., 2006; for more recent estimated origin). More generally, Fig. 5 shows the \(a\) or \(f\) values needed to get various \(r_o\) ratios as a function of \(\beta\). As \(\beta\) decreases, that is, the DFEM becomes more leptokurtic, a more marked reduction in \(N_e\) (lower \(a\)) or more ancient transitions (higher \(f\)) are necessary to detect differences in \(\omega\) between SELF and OUT branches. This is because most mutations have very small effects and behave almost neutrally in both selfing and outcrossing species.

**Discussion**

**Weak signature of relaxed selection in selfers**

We compiled 19 chloroplastic data sets (16 of which provided enough information to infer realistic \(\omega\) values) in 13 angiosperm groups. Only a few (one to three, depending on the analysis) data sets showed significant evidence of elevated \(\omega\) in SELF compared to OUT branches. By combining data sets, a general trend emerged, but it was partly due to the fact that SELF
branched were mostly terminal. Overall, we found weak statistical support for increased $\omega$ in SELF branches. However, the simulation results suggested that these results were compatible with true weak-to-moderate differences in $\omega$ (Fig. 4).

Previous studies using similar divergence analyses also found a nil or weak signature of relaxed selection in selfers, whereas relaxed selection seemed to be more easily detected using polymorphism data (see references in Table 1). This could simply be due to the fact that many other factors can mask the relationship between mating systems and $\omega$ in large-scale phylogenetic analyses. Here, we also used proxies of selfing rates, such as self-compatibility and homostyly. Self-compatible species are more likely on the way to high selfing but could transiently have intermediate selfing rates (Igic et al., 2008; Igic & Busch, 2013). Under mixed mating, reduced selection efficacy is not expected to be as strong, unless bottleneck effects predominate. However, the signature of relaxed selection was not lower for these data sets than for data sets with ‘true’ selfing species. As we analysed chloroplastic genes, the expected reduction in $N_e$ could be a bit lower than for nuclear genes in species with few chromosomes and/or a short nuclear genetic map, as discussed in the introduction. This could also contribute to the weak signature of relaxed selection. For instance, the selfer Arabidopsis thaliana has a small genome with a rather short genetic map (five chromosomes of average length 100 cM, which roughly corresponds to the thin black lines in Fig. 1). In this species, the reduction in effective population size as compared to the outcrosser A. lyrata was found to be less severe in chloroplastic than in nuclear genes, in agreement with lower background selection affecting chloroplast (Wright et al., 2008). Chromosome numbers are higher in most angiosperm species (including those studied here), and the differences between chloroplast and nuclear genes could be rather low. However, the difference also depends on the rate and effect of deleterious mutations (Fig. 1). A better characterization of both genetic maps and the distribution of mutational effects would be needed to evaluate the respective impact of background selection on chloroplastic and nuclear genes. It would also be interesting to confirm (or not) our results using nuclear data.

The lack of any sign of relaxed selection has also been interpreted as evidence of a recent origin of selfing lineages. Our results seem to support this, but we also found that this interpretation closely depends on the underlying distribution of fitness effects of mutations (DFEM) affecting the genes used in the Dn/Ds analyses (Eqns 3 and 4). Figure 5 shows that the interpretation holds only if this distribution is not too leptokurtic ($\beta$ not too small). We do not know the $\beta$ parameters for the matK and rbcL genes in the different species groups we used. However, several genome-wide estimates based on sequence polymorphism data suggest that $\beta$ is rather small (e.g. 0.23 in humans, see Eyre-Walker et al., 2006; between 0.08 and 0.21 in several plant species T. Gossmann pers. comm. and Gossmann et al., 2010). With such low $\beta$ values, a marked $N_e$ reduction in selfers (small $\omega$) over a long time (if close to 1 in eqn 4) will only lead to slight differences in $\omega$, as observed in the data sets. A more ancient selfing origin, that is, from the beginning of the branches leading to current selfing species, would be in line with the recent finding that a change in mating systems more frequently occurs in association with speciation events (‘cladogenetic’ mode of change) than within lineages (‘anagenetic’ mode) (Goldberg & Igic, 2012).

Several biological factors have been proposed to explain the weaker signature of selfing than asexuality in $\omega$ analyses (Glémin & Galtier, 2012). Differences in the shape of the DFEM of genes used in the different analyses (mostly mitochondrial in asexual animals vs. mostly nuclear or chloroplastic in selfing plants) could also play a role. If possible, choosing genes with high $\beta$ should facilitate the detection of relaxed selection caused by a reduction in population size.

**Selfing evolution implications**

Based on the assumptions underlying eqns (3) and (4), our results are thus compatible with either small $\beta$ and possibly a strong reduction in $N_e$ over a long time or a higher $\beta$ and a recent origin of selfing lineages. But does $\beta$ matter? Considering the risk of population extinction due to the accumulation of deleterious mutations, and assuming that mutation effects have a gamma distribution, Lande (1994) showed that the mean time to extinction is approximately proportional to $N_e^{\beta+1}$ for $\gamma > 1$ (his equation 10). A reduction in $N_e$ by an $\alpha$ factor thus reduces the mean time to extinction by an $\alpha^{\beta+1}$ factor. If $\beta$ is weak, a reduction in $N_e$ reduces the mean time to extinction almost linearly. For higher $\beta$, a reduction in $N_e$ has a more drastic effect on the mean time to extinction. For example, for $\beta = 1$ (i.e. an exponential distribution of deleterious effects), reducing the effective population size by two-fold reduces the mean time to extinction by four-fold. If $\beta$ is small, our results are compatible with a marked reduction in $N_e$ in selfers, but the impact on extinction through mutational meltdown may be rather weak. If $\beta$ is higher, deleterious mutations may have a stronger impact on species extinction, but they likely accumulate for a much shorter time. Overall, although the reasons are maybe more complex than previously thought (Glémin & Galtier, 2012), this suggests that deleterious mutations may not be the main cause of extinction in selfing species.

However, understanding the causes of higher extinction rates in selfing species is still a difficult task, and
the possible role of deleterious mutations should be tested more quantitatively. Our results suggest that better characterization of DFEM and mutation rates in selfing and outcrossing species could help in addressing this issue. Moreover, theoretical work is still needed to test whether the accumulation of deleterious mutations could be sufficient to cause species extinction but without leaving a strong molecular signature. Finally, it would also be crucial to clarify the demographic consequences of the mutation load in an ecological context (Agrawal & Whitlock, 2012) to resolve this question.

Acknowledgements

We thank Stephen Wright and an anonymous reviewer for their helpful comments and suggestions on a previous version of the manuscript. We declare no conflict of interest. This publication is the contribution ISEM 2014-013 of the Institut des Sciences de l’Évolution de Montpellier (UMR 5554 – CNRS). This work was supported by the French Centre National de la Recherche Scientifique and Agence Nationale de la Recherche (ANR-11-BSV7-013-03).

Authors’ contribution

SG designed the project. AM and SG built the data sets and analysed data. SG developed theoretical predictions. SG and AM wrote the manuscript.

References


Appendix 1  
Effect of background selection on nuclear and chloroplastic genes

In this JEB issue, Kamran-Disfani and Agrawal obtained an expression for the reduction in $N_e$ caused by background selection in a partially fertilizing population, which is more accurate than the approximated one obtained previously (Cutler & Payseur, 2003; Glémin, 2007; Glémin & Ronfort, 2013). Consider a chromosome of size $C$ and a focal gene located at position $z$ and assuming the same deleterious mutation rates, $u$, dominance, $h$, and selection effect, $s$, for all selected genes on the chromosome, they found that:

$$
\rho_{\text{linked}}(F, C, z) = \exp \left[ - \int_0^z \frac{u(hF - hF^2)}{2((hF - hF^2)(1 - F)R(x))} \, dx \right]
\rho_{\text{unlinked}}(F, G - C) = \exp \left[ - \int_0^{C(1 - z)} \frac{u(hF - hF^2)}{2((hF - hF^2)(1 - F)R(x))} \, dx \right]
$$

where $F$ is Wright’s fixation index and $R(x) = \frac{1}{2}(1 - \exp[-2rx])$ is Haldane’s mapping function, with $r$ being the recombination rate between two adjacent sites. Kamran-Disfani and Agrawal gave the exact analytical expression for (A1.1) in supplementary material, but the equation is unwieldy and not reproduced here. This result, which is based on the findings of Hudson & Kaplan (1995), can be easily extended to the effect of deleterious alleles segregating at loci not physically linked to the focal gene, simply using $R(x) = 1/2$ and integrating between 0 and G – C, where $G$ is the total genome size. G – C thus corresponds to the other chromosome(s) of the genome.

$$
\rho_{\text{unlinked}}(F, G - C) = \exp \left[ - \int_0^{G-C} \frac{u(hF - hF^2)}{2((hF - hF^2)(1 - F/2)^2)} \, dx \right]
\rho_{\text{unlinked}}(F, G - C) = \exp \left[ - \int_0^{C(1 - z)} \frac{u(hF - hF^2)}{2((hF - hF^2)(1 - F/2)^2)} \, dx \right]
$$

(A1.2)

The total effect of background selection is thus as follows:

$$
\rho_{\text{nuc}}(F, G, C, z) = \rho_{\text{unlinked}}(F, G - C)\rho_{\text{linked}}(F, C, z)
$$

and the $z$ term is given by as follows:

$$
\sigma_{\text{nuc}}(F, G, C, z) = \rho_{\text{nuc}}(F, G, C, z) / \rho_{\text{nuc}}(0, G, C, z)
$$

(A1.4)

For a chloroplastic gene, we can neglect the effect of other linked chloroplastic genes, so we simply have as follows:

$$
\rho_{\text{chlo}}(F, G) = \rho_{\text{unlinked}}(F, G)
$$

(A1.5)

and

$$
\sigma_{\text{chlo}}(F, G) = \rho_{\text{chlo}}(F, G) / \rho_{\text{chlo}}(0, G)
$$

(A1.6)

We can show that $\lim_{Cr \to \infty} \rho_{\text{linked}}(F, C, z) = \rho_{\text{unlinked}}(F, C)$, so when the chromosome genetic map is large, $Cr \gg 1$, $\rho_{\text{nuc}}(F, G, C, z)$ tends towards $\rho_{\text{unlinked}}(F, G) = \rho_{\text{chlo}}(F, G)$. Similarly, when $G > C$, $\rho_{\text{nuc}}(F, G, C, z) > \rho_{\text{unlinked}}(F, G)$ and $\rho_{\text{chlo}}(F, G)$. The additional reduction in $N_e$ due to background selection in selfing species, $\sigma(F)$, is thus similar for chloroplastic and nuclear genes except for genomes with few chromosomes and a short genetic map. This is illustrated in Fig. 1.
Appendix 2
Derivation of eqns (3) and (4)

Assuming a gamma distribution of deleterious effects of mutations with mean $\gamma = Ns$ and shape $\beta$, Welch et al. (2008) showed that $\omega$ can be well approximated by:

$$\omega(\gamma, \beta) \approx \frac{1}{\zeta(1+\beta)} \frac{1}{\gamma^{1+\beta} \zeta(1+\beta)} \zeta(1+\beta)$$  \hspace{1cm} (A2.1)

where $\zeta$ is the Riemann zeta function (Abramowitz & Stegun, 1970). Here, $\gamma = Ns$ and not $\gamma = 4Ns$ as in Welch et al. (2008) because the chloroplastic genome is haploid and uniparentally transmitted. If we assume that the shape $\beta$ is the same between outcrossing and selfing species, we simply have as follows:

$$r_o = \frac{\omega_{\text{SELF}}}{\omega_{\text{OUT}}} \approx \frac{\omega(\gamma, \beta)}{\omega(\gamma, \beta)} = r_o^{\gamma, \beta}$$  \hspace{1cm} (A2.2)

hence eqns (3) and (4). For a fixed $r_o$ ratio, we thus have the following relationship between $x$ and $\beta$ as plotted in Fig. 5a:

$$x = r_o^{-1/\beta}$$  \hspace{1cm} (A2.3)

and between $f$ and $\beta$, as plotted in Fig. 5b:

$$f = \frac{\beta(1-r_o)}{\beta^2 - 1}$$  \hspace{1cm} (A2.4)

Supporting information

Additional Supporting Information may be found in the online version of this article:

**Table S1** List of species and Genbank accessions.

**Table S2** Details results of the comparison between M7 and M8 models in codeml to detect codons under positive selection.

**Table S3** Details of simulation results.

Data deposited at Dryad: doi:10.5061/dryad.618fv

Received 1 October 2013; revised 10 February 2014; accepted 12 February 2014