Molecular evolution of freshwater snails with contrasting mating systems

Concetta C. Burgarella, Philippe Gayral, Marion Ballenghien, Aurelien Bernard, Patrice David, Philippe Jarne, Ana Corrêa, Sylvie Hurtrez-Boussès, Juan Escobar, Nicolas Galtier, et al.

Molecular Evolution of Freshwater Snails with Contrasting Mating Systems

Concetta Burgarella,*[1] Philippe Gayral,†[,1] Marion Ballenghien,[1] Aurélien Bernard,[1] Patrice David,[2] Philippe Jarne,[2] Ana Correa,[3] Sylvie Hurtrez-Boussès,[3] Juan Escobar,‡[,1] Nicolas Galtier,[1] and Sylvain Glémin[1]

[1]Institut des Sciences de l'Evolution, UMR, CNRS 5554, Université Montpellier II, Montpellier, France

[2]CEFE/CNRS Campus du CNRS 1919, Montpellier, France

[3]MIVEGEC (Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution, Contrôle), UMR (UM1-UM2-CNRS 5290-IRD224), IRD, Montpellier, France

†Present address: Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS, Université François-Rabelais, Tours, France

‡Present address: Vidarium—Nutrition, Health and Wellness Research Center, Grupo Nutresa, Medellín, Colombia

***Corresponding author:** E-mail: concetta.burgarella@gmail.com.

Associate editor: Stephen Wright

Abstract

Because mating systems affect population genetics and ecology, they are expected to impact the molecular evolution of species. Self-fertilizing species experience reduced effective population size, recombination rates, and heterozygosity, which in turn should decrease the efficacy of natural selection, both adaptive and purifying, and the strength of meiotic drive processes such as GC-biased gene conversion. The empirical evidence is only partly congruent with these predictions, depending on the analyzed species, some, but not all, of the expected effects have been observed. One possible reason is that self-fertilization is an evolutionary dead-end, so that most current selfers recently evolved self-fertilization, and their genome has not yet been strongly impacted by selfing. Here, we investigate the molecular evolution of two groups of freshwater snails in which mating systems have likely been stable for several millions of years. Analyzing coding sequence polymorphism, divergence, and expression levels, we report a strongly reduced genetic diversity, decreased efficacy of purifying selection, slower rate of adaptive evolution, and weakened codon usage bias/GC-biased gene conversion in the selfer *Galba* compared with the outcrosser *Physa*, in full agreement with theoretical expectations. Our results demonstrate that self-fertilization, when effective in the long run, is a major driver of population genomic and molecular evolutionary processes. Despite the genomic effects of selfing, *Galba truncatula* seems to escape the demographic consequences of the genetic load. We suggest that the particular ecology of the species may buffer the negative consequences of selfing, shedding new light on the dead-end hypothesis.

Key words: **mating systems, molecular evolution, freshwater snails, selfing, selection, base composition.**

Introduction

Life-history and ecological traits are expected to influence genome organization and evolution through their effects on population genetic processes (Lynch 2007). In particular, mating systems affect fundamental population genetic parameters, such as effective population size, recombination rates, and the efficacy of natural selection. Therefore, they potentially influence patterns of polymorphism, rates of molecular evolution, and genomic base composition (Jarne 1995; Charlesworth and Wright 2001; Wright et al. 2008; Glémin and Galtier 2012). Selfing is a widely distributed reproductive mode, fairly common in various families of plants and animals (Vogler and Kalisz 2001; Jarne and Auld 2006). Compared with outcrossing, selfing is expected to reduce the effective population size, N_e, because of nonindependent gamete sampling and prevalence of genome-wide homozygosity (Pollak 1987; Nordborg 2000). All other things being equal, $N_e = N/(1 + F_{IS})$, where F_{IS} is the Wright's inbreeding coefficient, and N the effective size under panmixia, so that N_e is halved in case of complete self-fertilization (when $F_{IS} = 1$). N_e is expected to be reduced further 1) by hitchhiking effects, such as background selection (Charlesworth et al. 1993; Kamran-Disfani and Agrawal 2014), because selfing also reduces effective recombination rates (Nordborg 2000); and 2) by bottleneck effects, which are thought to be more frequent in selfers because a single or a few individuals can create a new population (Schoen and Brown 1991; Jarne 1995). At the species levels, stronger population genetic structure (higher F_{ST}) is expected in selfers than in outcrossers, because of both higher local genetic drift and reduced gene flow through the male function (for organisms that disperse during the male haploid phase). High F_{ST} can increase global N_e (Whitlock and Barton 1997); however, when extinction–recolonization dynamics occur, selfing also reduces N_e at the species level (Ingvarsson 2002). This should be reinforced by the local reduction of N_e through hitchhiking effects.

As levels of neutral genetic diversity are proportional to $N_e\mu$, where μ is the mutation rate, lower polymorphism levels are expected in selfers, both at the population and the species level (table 1). Along the same lines, the efficacy

Table 1. Effects of Mating System on Genomic Features According to Predictions and Observed in This Study.

	Predicted		Observed
	Outcrossing	Selfing	
Polymorphism	+	−	**Yes**
Populations structure (F_{IS}, F_{ST})	−	+	**Yes**
Selection efficacy (purifying selection, adaptive selection, selection on codon usage)	+	−	**Yes**
Sexual conflicts (rapid evolution of male gametes)	+	−	**No**
gBGC	+	−	**Yes**

NOTE.—F_{IS}, heterozygote deficiency due to inbreeding; F_{ST}, heterozygote deficiency due to population structure.

of selection, which depends on $N_e s$ (where s is the selection coefficient; Kimura 1983), should be lower in selfing species (table 1). If a large fraction of mutations are weakly deleterious, as supported empirically (Eyre-Walker and Keightley 2007), we thus expect higher nonsynonymous versus synonymous polymorphism (π_N/π_S) and divergence (d_N/d_S) ratios in selfing than in outcrossing lineages, reflecting the accumulation and fixation of deleterious mutations (Glémin 2007). For moderate to high selfing rates, purifying selection is also predicted to be reduced in subdivided populations (Roze and Rousset 2004). However, strongly deleterious and partly recessive mutations should be more easily purged in selfing species, somewhat leveling the above-mentioned effect (Glémin 2003). On the other hand, fixation of advantageous alleles should be reduced in selfers (at least when they are not too recessive), lowering the rate of adaptive substitutions (Glémin 2007). Similarly, bias in codon usage is expected to be weaker in selfers due to reduced efficacy of selection on translational speed/accuracy. Finally, mating systems might also affect the genome base composition through GC-biased gene conversion (gBGC, hereafter), a recombination-associated mechanism favoring G and C bases during mismatch repair occurring at meiosis (Marais 2003). gBGC is expected to be null or very low in selfers (table 1) as it occurs only at heterozygous sites (Marais et al. 2004). Beyond genome-wide effects, specific genes or gene categories should also exhibit strong contrasting patterns between selfers and outcrossers (table 1). For instance, direct selection intensity (s not only $N_e s$) should be relaxed in selfers at some genes involved in reproduction (especially male expressed genes; Slotte et al. 2010) or genomic conflicts (Kawabe et al. 2007; Spillane et al. 2007).

Published empirical studies have addressed almost exclusively plant species (reviewed in Glémin and Galtier 2012; Glémin and Muyle 2014). They mainly agree with theoretical predictions regarding nucleotide diversity, that is, lower π_S and higher π_N/π_S in selfers (Glémin and Muyle 2014). However, patterns of divergence (d_N/d_S and adaptive substitution rates) and base composition comparisons are less clear-cut (e.g., Wright et al. 2002; Cutter et al. 2008; Haudry et al. 2008; Escobar et al. 2010; but see Qiu et al. 2011). Several reasons have been proposed to explain this. First, selfing could be of too recent origin, so that its impact would only be detectable at a short time scale (i.e., polymorphism; Wright et al. 2002; Bechsgaard et al. 2006; Cutter et al. 2008; Escobar et al. 2010): If selfing has evolved recently relative to the species divergence time, its effects concern only a small fraction of the overall divergence, which would therefore not reflect the effect of extant mating systems. Second, if nonsynonymous substitutions were predominantly adaptive, this would reverse the prediction of an increased d_N/d_S in selfers (Lanfear et al. 2014). Finally, it has been shown that gBGC can counteract selection and lead to the fixation of weakly deleterious GC alleles in highly recombining regions (Galtier et al. 2009; Glémin 2010), which could contribute to increase d_N/d_S in outcrossing species (Haudry et al. 2008).

In this article, we assess the main predictions on the effects of mating system on molecular evolution with empirical data, trying to control for the main issues mentioned above. To disentangle the signatures of purifying and positive selection, we analyzed polymorphism and divergence patterns and addressed the possible interaction between gBGC and selection from transcriptomic data in two groups of hermaphroditic freshwater snails (pulmonate gastropods) with similar biology and ecology but contrasting mating systems—outcrossing versus selfing. This group of snails is the best studied animal group with regards to mating system (Escobar et al. 2011, and references therein). Estimates of selfing rate, inbreeding depression, and waiting time to selfing are available for many species, based on genetic and reproductive biology data. Within the group, predominantly outcrossing species show high inbreeding depression and long waiting times. Among them, we choose *Physa acuta* and *Physa gyrina* (as outgroup), with estimated outcrossing rates greater than 90% (Henry et al. 2005; Escobar et al. 2011) and greater than 70% (Escobar et al. 2011), respectively. Predominantly selfing species, instead, show low inbreeding depression and no waiting time. Among them, we selected *Galba truncatula*, with estimated outcrossing rates less than 2% (Trouvé et al. 2003; Meunier et al. 2004), and *Galba cubensis* (outgroup) that shows extremely low heterozygosity and polymorphism at 16 microsatellite markers (Lounnas M, Vazquez AA, Hurtrez-Boussés S, unpublished data) and presents morphologic and behavior characteristics typical of selfing species (Correa et al. 2011).

Our species choice is particularly well suited to our purpose for several reasons. First, the two focal species, *P. acuta* and *G. truncatula*, show similar key life-history traits, ecological features, and distribution, so that we would not expect significant difference in the species-level effective population sizes between them beyond the effect of mating system. Actually, they show similar reproductive traits, such as life span, number of generations per year (Heppleston 1972; Li et al. 2014), adult and egg size (Tsitrone et al. 2003; Bargues et al. 2011; Awdziejczyk and Jaeckle 2012). In particular, the propagule/adult size ratio, which has been shown to be a major determinant of the genetic diversity at the species level in animals (Romiguier et al. 2014), is very similar between

Physa and *Galba*. Both species occupy permanent and temporal freshwater environments (Henry et al. 2005; Trouvé et al. 2005), have cosmopolitan distribution (Dillon et al. 2002; Correa et al. 2010; Bousset et al. 2014) and demonstrate high colonizing potential (Trouvé et al. 2005; Bousset et al. 2014). Outgroups *P. gyrina* and *G. cubensis* have a narrower but comparable distribution, North America and the Neotropics, respectively (Dillon 2000; Correa et al. 2011). Second, it is very likely that the studied species pairs have shared the same mating system since their split, such that in each pair the whole divergence would have occurred under the same mating system. Each species pair is included in a well-supported clade (Wethington and Lydeard 2007; Correa et al. 2010; Dayrat et al. 2011). In the Galba clade (fig. 1), all the species studied so far based on breeding experiments and/or molecular markers are predominantly selfing (see references on fig. 1). The evolution of selfing therefore likely predated the common ancestor to extant species in this clade. Consequently, the divergence and the age of selfing of species analyzed here are probably far more ancient than those estimated for the plant species studied up to now, enough for selfing to leave a detectable signature. Based on the phylogeny of Correa et al. (2010), we estimate that the age of the divergence of *Galba* species would be at least 20 My (using fossil calibration, fig. 1) or approximately 23 My (using the substitution rate of 0.22×10^{-8} per site per year estimated in bivalve mollusks by Coleman and Vacquier [2014]). The estimated ages of divergence between focal and outgroup in previously studied plant species are comparatively much more recent—for instance 200,000 years in *Capsella rubella* (Slotte et al. 2013), and less than 1 My in *Arabidopis thaliana* (Charlesworth and Vekemans 2005; Bechsgaard et al. 2006).

Results

To compare the effect of mating system on genomic patterns, we used transcriptomic data in the two pairs of closely related species: The selfers *Galba truncatula* (focal species) and *G. cubensis* (outgroup), and the outcrossers *Physa acuta* (focal) and *P. gyrina* (outgroup). Table 2 summarizes the main data characteristics and provides the main population genomic statistics, from initial raw contigs to final analyzed genes, for all species and species pairs. As expected, the number of usable genes was lower for the focal–outgroup species pairs (1,614 and 1,533 for *G. truncatula*–*G. cubensis* and *P. acuta*–*P. gyrina*, respectively) than for single focal species (3,976 and 3,015 for *G. truncatula* and *P. acuta*, respectively). Polymorphism estimates within focal species varied slightly depending on whether all genes or only genes shared with the outgroups were analyzed, reflecting a bias toward more conserved genes among the second gene set, consistent with Gayral et al. (2013). Among retained genes, those with a hit to known proteins (*e* value < 0.001) were 83% and 88% for *G. truncatula* and for *P. acuta*, respectively (76% and 78% of the first hits were assigned to Mollusca, respectively).

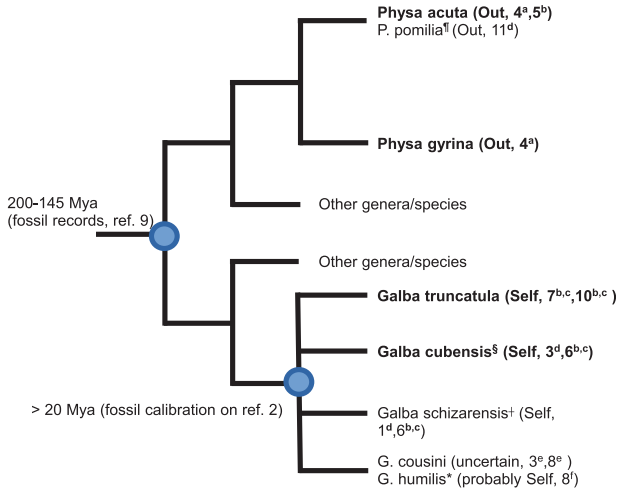

Fig. 1. Phylogenetic relationships among the outcrossing species *Physa acuta* (focal) and *P. gyrina* (outgroup) and the selfing *Galba truncatula* (focal) and *G. cubensis* (outgroup). Modified from references 2 and 12. In bold the species analyzed in this study. Out, predominantly outcrossing; Self, predominantly selfing; NA, no data available. References: 1. Bargues et al. 2011; 2. Correa et al. 2010; 3. Correa et al. 2011; 4. Escobar et al. 2011; 5. Henry et al. 2005; 6. Louna M, Vazquez AA, Hurtrez-Boussès S, unpublished data; 7. Meunier et al. 2004; 8. Poitier JP, unpublished data; 9. Taylor 1988; 10. Trouvé et al. 2003; 11. Wethington and Dillon 1997; 12. Wethington and Lydeard 2007. ¶*Physa hendersoni* from 12 is synonymous of *P. pomilia*; §*G. viatrix*, *G. neotropica*, and *Fossaria bulimoides* from reference 2 are synonyms of *G. cubensis*; †same species as *Galba* sp. of reference 2; **F. obrussa* from reference 2 is synonym of *G. humilis*. Type of evidence regarding the mating system: [a]laboratory measures of waiting time before selfing; [b]microsatellite data on progeny arrays; [c]microsatellite-based estimate of population inbreeding coefficient; [d]breeding experiments; [e]shell morphology and anatomy of reproductive organs departing from the rest of the clade; [f]anatomy of reproductive organs.

Mating Systems and Polymorphism Patterns

We estimated *F*, the deviation from Hardy–Weinberg expectation. This is equivalent to F_{IT} and mainly corresponds to F_{IS} for selfing species and to F_{ST} for outcrossing ones. First, we assumed panmixia ($F = 0$) in the genotype calling and paralog filtering procedure for all species (see Materials and Methods). We found $F = 0.91$ for *G. truncatula*, confirming its highly inbred status. This value was then used for a second run of single nucleotide polymorphism (SNP) and genotype calling in *G. truncatula* and *G. cubensis*. After this second run, *F* was 0.98 for *G. truncatula*. All the results presented below for *G. truncatula* come from this latter analysis. The observed $F = 0.23$ for *P. acuta* likely reflects the geographic structure of the sample. Because this value is rather low, we kept the value obtained with *F* fixed to 0 as done in Romiguier et al. (2014). However, we verified that using $F = 0.23$ gave very similar results (supplementary table S1, Supplementary Material online).

Physa acuta showed a 5-fold higher diversity than *G. truncatula* (mean $\pi_S = 0.030$ and 0.006, respectively; table 2), consistent with the expectations of a higher N_e associated with the outcrossing mating system (table 1). To get an insight into the recent demographic history of the two species, we used

Table 2. Characteristics of the Data Sets, Coding Sequence Polymorphism, Divergence Patterns, and Base Composition in the Selfing (*Galba truncatula* and *G. cubensis*) and Outcrossing Species (*Physa acuta* and *P. gyrina*).

Species	*G. truncatula*	*G. cubensis*		*P. acuta*	*P. gyrina*	
Type of analysis	Focal	Outgroup	Focal and Outgroup	Focal	Outgroup	Focal and Outgroup
No. of individuals	10	2		9	2	
10^6 reads (total)	24.5	2.2		12.9	2.9	
No. of raw contigs	91,835	36,388		127,054	52,583	
Filtered contigs	13,231	11,857		12,153	19,093	
No. of ORFs > 200 bp	9,547	6,422		8,009	8,887	
ORF length (bp): median (95% quantiles)	723 (210, 3,028)	417 (204, 1,654.4)		576 (210, 2,297.4)	351 (204, 1,457.6)	
No. of retained coding sequences[a]	3,976	1,969	1,614[b]	3,015	2,340	1,533[c]
Mean length (bp) of coding sequences (95% quantiles)	664.9 (138, 1,925.2)	799.6 (217, 1,981)	585.7 (42, 1,656)	480.9 (69, 1,572)	680 (208, 1,999)	450.9 (36, 1,489.2)
Mean no. of complete, biallelic sites (95% quantiles)	103.21 (11, 397)	138.74 (12, 485.8)	142.42[b] (7, 455.7)	83.34 (11, 344.2)	115.93 (11, 448)	100.14[c] (6, 376.4)
No. of SNPs	13,002	838	7,316[b]	34,972	8,167	20,524[c]
π_S% (CI)	0.56 (0.54, 0.58)	0.2 (0.17, 0.24)	0.57[b] (0.54, 0.60)	3.03 (2.90, 3.17)	1.87 (1.76, 2.00)	2.88[3] (2.72, 3.04)
π_N% (CI)	0.14 (0.13, 0.15)	0.07 (0.06, 0.09)	0.12[b] (0.11, 0.13)	0.23 (0.21, 0.25)	0.21 (0.19, 0.23)	0.15[c] (0.14, 0.17)
π_N/π_S (CI)	0.24 (0.23, 0.26)	0.37 (0.31, 0.43)	0.21[b] (0.19, 0.23)	0.08 (0.07, 0.08)	0.11 (0.10, 0.12)	0.05[c] (0.05, 0.06)
d_S% (CI)			10.60[b] (0.10, 0.11)			5.62[c] (5.38, 5.88)
d_N% (CI)			1.33[b] (1.23, 1.44)			0.59[c] (0.53, 0.65)
d_N/d_S (CI)			0.13[b] (0.12, 0.13)			0.11[c] (0.09, 0.12)
Mean GC content (95% quantiles)	0.44 (0.36, 0.54)			0.44 (0.35, 0.56)		
Mean GC3 content (95% quantiles)	0.43 (0.27, 0.66)			0.43 (0.24, 0.74)		

[a]After UTR sequence removal and reads2snp aligned site cleaning.
[b]Calculated on common genes between *G. truncatula* and *G. cubensis*.
[c]Calculated on common genes between *P. acuta* and *P. gyrina*.

the synonymous derived allele frequency (DAF) spectra to fit two simple demographic scenarios. For both species, the model with a population size change fit much better the observed DAF spectra than the null model with constant population size and both species showed signature of demographic expansion, especially *G. truncatula* (supplementary material S2, Supplementary Material online).

Mating Systems and the Intensity of Purifying Selection

The global strength of purifying selection was evaluated through the ratio of nonsynonymous to synonymous polymorphisms, π_N/π_S. It was considerably lower in *P. acuta* (mean $\pi_N/\pi_S = 0.08$) than in *G. truncatula* (mean $\pi_N/\pi_S = 0.24$) (table 2). Because expression level is usually found to affect the intensity of purifying selection (Drummond et al. 2005; Park et al. 2013; Nabholz et al. 2014), we verified that the difference was not due to a difference in expression levels between the two species. Though π_N/π_S decreases with expression level in both species, the difference between species is observed for any class of expression. In the analysis of covariance (ANCOVA), species explains more than 67% of the variation in π_N/π_S (P value $< 10^{-15}$; fig. 2) whereas expression level only explains 1.3% (P value < 0.01) and the interaction is not significant. For each species pair, we used the folded synonymous and nonsynonymous site frequency spectra (SFS) to estimate the genome-wide distribution of fitness effects (DFE) of nonsynonymous mutations (Eyre-Walker et al. 2006; Eyre-Walker and Keightley 2009). We also found striking differences between the two species (fig. 3). Consistent with its lower N_e, we inferred a higher proportion of nonsynonymous nearly neutral mutations in *G. truncatula* than in *P. acuta* (17% vs. 5% for $0 \leq 4N_es \leq 1$) and a lower proportion of highly deleterious mutations (67% vs. 85% for $4N_es > 100$). Finally, the d_N/d_S ratio was only slightly lower in outcrossers than in selfers (0.11 vs. 0.13; table 2) although the difference was significant (Wilcoxon signed rank test $P = 0.0017$), suggesting that either the intensity of selection was more similar between the two species in the past than recently, or that other processes obscure the difference between species (see below). The outgroup species *P. gyrina* and *G. cubensis* showed polymorphism patterns very similar to *P. acuta* and *G. truncatula*, respectively, even though estimates relied on just two individuals in these species (table 2). This underpins the idea that the observed

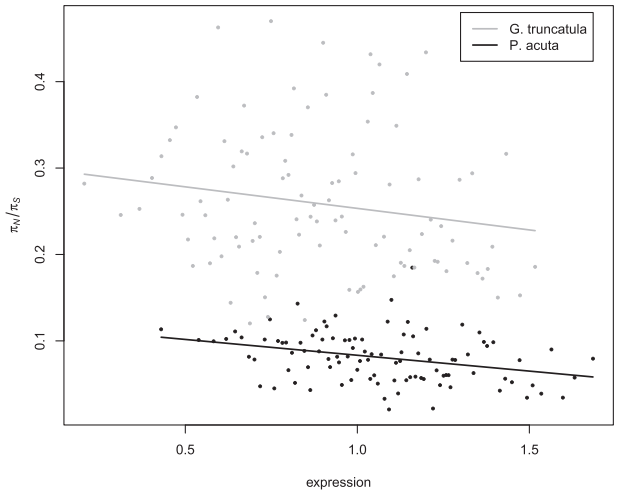

Fig. 2. Correlation between π_N/π_S and expression in *Galba truncatula* and *Physa acuta*. *X* axis: Log transformed coverage used as proxy for expression. Average coverage per gene was calculated by summing the length of matching reads and dividing the result by the cDNA length and the total number of reads sequenced. *Y* axis: Mean π_N/π_S for 100 coverage classes built as to minimize the difference in the number of polymorphisms between classes.

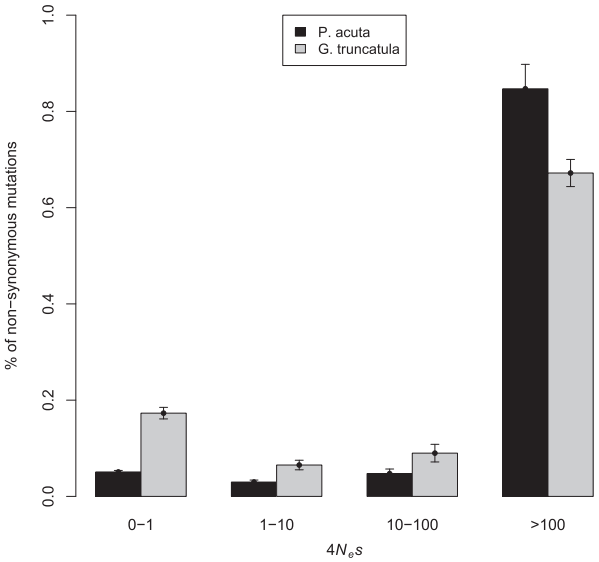

Fig. 3. Distribution of the fitness effect of nonsynonymous mutations falling in different N_es ranges in the outcrossing species *Physa acuta* and the selfing species *Galba truncatula*.

differences between selfers and outcrossers are not species-specific but likely driven by the difference in mating system.

Mating Systems and Adaptive Evolution

We used the "second" method of Eyre-Walker and Keightley (2009) to estimate the proportion of adaptive nonsynonymous substitutions, α, during the divergence between the focal species and its outgroup. α was 0.54 between *P. acuta* and *P. gyrina*. In contrast, no adaptive substitutions were found for *G. truncatula* ($\alpha < 0$) (table 3). Splitting nonsynonymous substitutions in adaptive (ω_a) and nonadaptive (ω_d), we obtained a higher adaptive rate and a lower nonadaptive rate in *Physa* than in *Galba* (table 3), in agreement with theoretical expectations (table 1). Values of ω_a indicate that the lower α in *G. truncatula* is not simply due to a higher proportion of slightly deleterious mutations contributing to the d_N/d_S ratio but also to a lower proportion of positively driven substitutions.

According to these analyses, adaptive substitutions strongly increase the d_N/d_S ratio in *Physa* but not in *Galba*, thus providing a plausible explanation for the low difference in d_N/d_S between the two groups. However, this rationale implicitly assumes that N_e has remained constant during species divergence. Alternatively, the small difference in d_N/d_S between the two species pairs might reflect ancient, contingent fluctuations in N_e. In supplementary material S3, Supplementary Material online, we showed that the increase in d_N/d_S ratio in *P. acuta*, compared with what is expected given the estimated DFE, is very unlikely explained solely by low ancient N_e, in fact a more than 100-fold N_e reduction in the past would be necessary to explain the observed d_N/d_S ratio without adaptation.

We also examined the direction and strength of selection in individual genes by conducting the McDonald–Kreitman (MK) test (McDonald and Kreitman 1991) per gene and by estimating the statistic DoS (Direction of Selection; Stoletzki and Eyre-Walker 2011). For *P. acuta*, 68 genes (five after applying the false discovery rate, FDR) deviated from neutral expectations in the direction of positive selection at the 5% significance level. In *G. truncatula*, only one (zero after FDR correction) gene was significantly under positive selection. The statistic DoS gave a similar picture as the MK analysis, in agreement with a prevalence of positive selection in the outcrossing species. Over all loci, there was a significant shift toward positive values for *P. acuta* (fig. 4, Wilcoxon signed rank test *P* value $< 10^{-15}$) where 44% of genes had a positive DoS and 27% had a negative DoS (36% and 49%, respectively, for *G. truncatula*). We did not find any evidence of relaxed selection in genes potentially involved in reproduction or response to stress with respect to the rest of the genome, as it could be expected in the selfing species. However, the absence of differences among gene categories might just reflect the imprecise GO annotation due to the lack of any available reference genome close to *Galba* and *Physa* species.

Mating Systems and the Evolution of Base Composition

We observed a similar average GC content (GC = 0.44 and GC3 = 0.43) in outcrossers and selfers (table 3). Codon preferences, inferred as in Duret and Mouchiroud (1999), are also quite similar between species. We found 20/24 and 21/27 putative preferred codons with guanine or cytosine in third position in *G. truncatula* and in *P. acuta*, respectively (supplementary table S2, Supplementary Material online). These observations would point at the absence of difference in base composition between outcrossing and selfing genomes. However, current GC patterns are not the best predictor of forces acting on them, because base composition is seldom at

Table 3. Parameters of the Fitness Distribution of Nonsynonymous Mutations (β, mean S), Rate of Adaptive Evolution (α), and Adaptive Proportion of d_N/d_S (ω_a) for the Selfing (*Galba truncatula*) and Outcrossing (*Physa acuta*) Species Estimated with the Method DoFE (Eyre-Walker and Keightley 2009).

	P. acuta	G. truncatula
β (CI)	0.2 (0.181, 0.229)	0.139 (0.128, 0.168)
Mean S (CI)	2.42×10^6 (3.47×10^5, 6.54×10^6)	1.76×10^5 (2.24×10^4, 3.5×10^5)
α (CI)	0.54 (0.51, 0.57)	−0.19 (−0.27, −0.10)
ω_a (CI)	0.05 (0.049, 0.057)	−0.019 (−0.028, −0.010)
ω_d (CI)	0.05 (0.039, 0.058)	0.149 (0.132, 0.165)

NOTE.—β, the shape parameter of the gamma distribution assumed for the estimation of S.

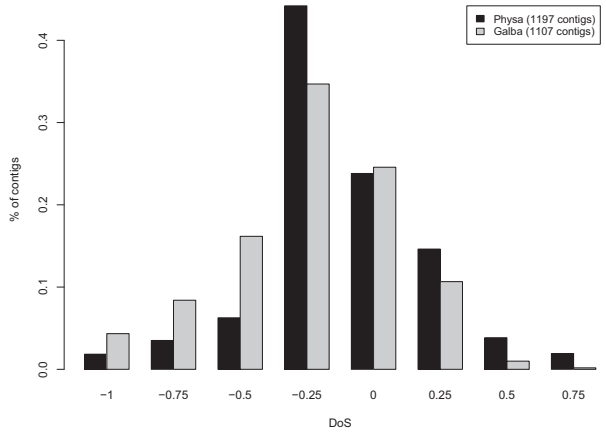

FIG. 4. Distribution of the DoS statistic across genes in the outcrossing *Physa acuta* and the selfing *Galba truncatula*.

equilibrium and it is usually more informative to look at base composition dynamics (e.g., Duret and Arndt 2008).

To infer base composition dynamics, we thus computed DAF spectra using outgroups to polarize mutations. They showed that AT toward GC (W→S) and unpreferred to preferred (U→P) mutations tend to segregate at higher frequencies than S→W and P→U alleles in both species (supplementary fig. S1, Supplementary Material online). However, SNP polarization errors can give spurious signature of selection or gBGC (Hernandez et al. 2007). Error rate is also expected to increase with divergence and may affect SFS more strongly in *G. truncatula* than in *P. acuta*, because divergence is twice as high in the *G. truncatula*–*G. cubensis* pair as in the *P. acuta*–*P. gyrina* pair (table 2). To circumvent this problem and to uncover the potential difference between the outcrossing and the selfing species, we fitted a population genetics model to the spectra to quantify the strength of selection $S = 4N_e s$ on codon usage or of gBGC, $B = 4N_e b$, taking into account the potential polarization errors (Glémin et al. 2014). Results are summarized in table 4. In all cases, the model including polarization errors showed a significantly better fit to the data than the model without error. Results of this model showed a great difference between the outcrosser and the selfer, in agreement with predictions (table 1). We found significant selection on codon usage and/or gBGC in *P. acuta* ($S = 1.42$, $B = 1.50$), whereas no significant evidence of any of these processes was found in *G. truncatula* (polarization errors explained the excess of high frequency alleles in spectra and including gBGC or selection did not significantly improve the model). As previously observed in many species, we also found a mutational bias toward A/T or unpreferred codons in both species (i.e., $\lambda > 1$, table 4).

The comparison between observed values of GC content and frequency of preferred codon (FoP) and predicted values at equilibrium, GC* and FoP*, confirms that base composition is not at equilibrium in either of the two species. GC content is decreasing in *G. truncatula* (GC = 0.43 vs. GC* = 0.22, FoP = 0.35 vs. FoP* = 0.21) and increasing in *P. acuta* (GC = 0.44 vs. GC* = 0.64, FoP = 0.38 vs. FoP* = 0.60). These trends in GC evolution are coherent with the expected effect of mating system on GC content.

gBGC can interfere with selection, leading to spurious signature of relaxed or positive selection (Galtier et al. 2009; Ratnakumar et al. 2010). This could increase the d_N/d_S ratio in outcrossing lineages, as proposed by Haudry et al. (2008). To test the hypothesis of an effect of gBGC on selection we sorted genes by increasing GC3 content, grouped them in five classes and computed π_N/π_S, d_N/d_S and DoS for each class. We did not find any effect of GC-content on these statistics (supplementary figs. S2 and S3, Supplementary Material online).

Confirmation of Results on a Subset of Orthologous Genes

To verify that the observed differences between species are not due to functionally different sets of genes, we recalculated 1) all polymorphism statistics for a subset of orthologous genes between the two focal species and 2) all divergence statistics for a subset of orthologous genes between all four species. The magnitude and direction of the difference in polymorphism and divergence patterns between outcrossers and selfers also hold when only the subset of genes common to *P. acuta* and *G. truncatula* was analyzed (supplementary table S3 and fig. S4, Supplementary Material online). DFE showed the same patterns as for the whole set of genes (supplementary table S4, Supplementary Material online). α was smaller for *Physa* ($\alpha = 0.174$), as expected if there were more conserved regions among the orthologous genes, but this value was still much higher than for *Galba* ($\alpha < 0$; supplementary table S4, Supplementary Material online). DoS could be calculated for only 98 orthologous genes for both species. Distributions were not statistically different, but *Physa* had the most positive DoS values and *Galba* the most negative (supplementary fig. S5, Supplementary Material online).

Discussion

Lower Ne and Lower Selection Efficacy in Selfers

Selfing is expected to reduce effective population size and recombination rate, so that self-fertilizing species should be less polymorphic than outcrossers and less efficient at purging slightly deleterious alleles or fixing new advantageous

Table 4. Estimation of gBGC (*B*), Selection on Codon Usage (*S*), and Mutation Bias (λ) in the Selfing (*Galba truncatula*) and Outcrossing Species (*Physa acuta*).

			Sat_M	M0	M1	M0_E	M1_E	M0_E versus M0 (*P* value)	M1_E versus M0_E (*P* value)
gBGC	*G. truncatula*	Log*L*	−61.23	−93.89	−68.78	−65.03	−64.99	$<10^{-11}$	0.774
		Λ		2.28	3.31	3.54	3.53		
		B		0	1.07	0	−0.18		
	P. acuta	Log*L*	−112.71	−232.8	−116.24	−140.55	−115.93	$<10^{-39}$	$<10^{-11}$
		Λ		1.64	2.5	2.32	2.55		
		B		0	1.36	0	1.5		
Selection on codon usage	*G. truncatula*	Log*L*	−55.64	−67.48	−58.92	−57.64	−57.64	$<10^{-3}$	0.978
		Λ		2.59	3.51	3.68	3.68		
		S		0	0.89	0	−0.02		
	P. acuta	Log*L*	−109.24	−179.16	−112.29	−124.97	−112.34	$<10^{-22}$	$<10^{-6}$
		Λ		1.76	2.73	2.53	2.72		
		S		0	1.4	0	1.42		

NOTE.—Sat_M, saturated model; M0, neutral mutation–drift model; M1, gBGC/selection model; M0_E, neutral model with polarization errors; M1_E, gBGC/selection model with polarization errors. *P* values from a LRT with 3 df (M0_E vs. M0) and 1 df (M1_E vs. M0_E).

mutations (Charlesworth and Wright 2001; Glémin 2007). We tested these hypotheses using transcriptomic data of two pairs of freshwater snail species with contrasting mating systems. Differently from most previous studies, we chose two pairs of species that very likely share the same mating systems since their common ancestor. Moreover, compared with most plant species studied so far, selfing is likely much older in these freshwater snail species. By combining polymorphism and divergence information we got a comprehensive insight into the effects of drift and selection on genome evolution of selfers versus outcrossers. Overall, our results conform well to predictions (table 1).

The two focal species share several ecological and life-history traits. Both are widespread species and show evidences of population expansion, which is consistent with the known history of colonization of all continents by both species and with their invasive character (Meunier et al. 2001; Dillon et al. 2002; Correa et al. 2010; Bousset et al. 2014). Recently, Romiguier et al. (2014) have shown that life-history traits are important determinants of polymorphism levels in animals. Species with an "*r*-like" strategy (i.e., little parental investment as measured by the propagule/adult size ratio), such as *P. acuta*, harbor the greatest diversity. As *G. truncatula* shares similar life-history traits and demographic history with *P. acuta*, we expect comparable levels of polymorphism in the two species. Contrary to these predictions but in agreement with the effect of mating system, we found that selfing species harbor much lower polymorphism (π_S) than outcrossers. This can be explained not only by the genetic effect of selfing on N_e but also by the particular population dynamics associated with selfing. Because they are able to find a population from a single individual, selfing species are more prone to high rate of population turnover, which is expected to reduce genetic diversity both at the local and the species scale (Jarne 1995; Ingvarsson 2002). As *G. truncatula* occupies patchy and ephemeral freshwater habitats, its populations suffer drastic density fluctuations leading to local extinction–recolonization dynamics (Trouvé et al. 2005). A further line of evidence supporting the effect of mating system is that the outgroup species, which share the same mating system, show similar polymorphism patterns as their focal species while they are much less widespread. This supports that mating system is the main determinant of the patterns we are reporting, though we cannot exclude that other unknown factors also contribute to these differences.

Furthermore, we observed in selfers several signatures of relaxed selection. On one hand, π_N/π_S and DFE comparisons indicated more slightly deleterious nonsynonymous mutations segregating in selfers than in outcrossers. Recently, Arunkumar et al. (2015) showed that, in addition to relaxed selection, the purging effect of selfing can also be detected by an excess of strongly selected variants ($N_e s > 100$) in the DFE in selfers compared with outcrossers, corresponding to stronger selection against highly recessive and deleterious alleles. This is especially true during the transition from outcrossing to selfing and when the N_e reduction is not too severe. Here, we did not detect any purging effect of selfing in this study, in agreement with either an ancient transition to selfing, a strong reduction in N_e, and/or moderately recessive mutations. On the other hand, we detected much lower genome-wide (α, ω_a) and single gene (MK test, DoS) signal of molecular adaptive evolution in selfers than in outcrossers. Theory also predicts that genome base composition should be affected by the mating system, because selection on codon usage and gBGC should be less effective in selfers than in outcrossers (Marais et al. 2004). Indeed, our estimates point to an effect of selection and/or gBGC only in the outcrosser species (table 4). Therefore, all the differences we observed between *G. truncatula* and *P. acuta* fit with the expected effect of their contrasting mating systems. A reason of concern might be that analyses were performed with different gene sets between species pairs. If the two sets of genes were experiencing different levels of selective constraint, distinctive polymorphism patterns might appear irrespective of mating systems. Nevertheless, the differences between *G. truncatula* and *P. acuta* are qualitatively the same using exclusively orthologous loci, for which mutational biases and selection pressures are likely to be similar, adding further support to our results (supplementary material S1, Supplementary Material online).

Divergence versus Polymorphism Pattern

In contrast to polymorphism patterns, the d_N/d_S ratio is only slightly higher in the selfing than in the outcrossing species pair. This is similar to what was found in previous studies that failed to detect signatures of relaxed selection as inferred by d_N/d_S (e.g., Wright et al. 2002; Cutter et al. 2008; Haudry et al. 2008; Escobar et al. 2010; Glémin and Muyle 2014; but see Gioti et al. 2013), whereas studies based on polymorphism data most often supported the expected effect of selfing (e.g., Slotte et al. 2010, 2013; Qiu et al. 2011; Hazzouri et al. 2013). The tempo of mating system evolution is often invoked to explain this discrepancy (reviewed in Glémin and Galtier 2012). If selfing was of relatively recent origin, divergence patterns would mainly reflect the effects of the ancient outcrossing history. Frequent adaptation, increasing the d_N/d_S ratio in outcrossers (Haudry et al. 2008; Slotte et al. 2010), and interference with gBGC (Haudry et al. 2008) have also been proposed to explain the lack of concordance between mating systems and d_N/d_S patterns.

Here, we circumvented all these potential drawbacks. First, we selected two selfing species, *G. truncatula* and *G. cubensis*, that belong to a clade whose extant species are prevalently selfing, so that the evolution of selfing likely predated divergence. Indeed, as mentioned above, polymorphism patterns of the outgroup species (table 2) are highly congruent with those of the focal species, supporting that focal and outgroup species share similar evolutionary dynamics. Second, by combining polymorphism and divergence data we clearly showed that the modest difference in d_N/d_S between the two pairs of species is mainly explained by a large fraction of adaptive substitutions in outcrossers. We also found no relationship between GC-content and d_N/d_S, suggesting that potential interference between gBGC and selection did not affect our results. Overall, our results highlight the importance of combining polymorphism and divergence to disentangle the effect of positive and purifying selection. They also demonstrate that selfing impacts many aspects of molecular evolution when it persists during a sufficiently long period of time.

Consequences for the Evolution of Selfing

Evolution from outcrossing to selfing is a frequent evolutionary transition in hermaphroditic species (Stebbins 1957, 1974; Jarne and Charlesworth 1993; Jarne and Auld 2006). On the short term, selfing is thought to be favored because of the 50% advantage of gene transmission (Fisher 1941) and the reproductive assurance under low pollen/mate availability (Darwin 1876; Baker 1955, 1967). Here, reproductive assurance is probably the major ecological advantage that drove the evolution of selfing in *G. truncatula* (Trouvé et al. 2005). By ensuring reproduction when mate density is low (e.g., following bottlenecks and founder events), selfing likely favors the persistence of *G. truncatula* in temporary habitats and contributes to its colonizing ability. However, on the long term, because of the potential accumulation of deleterious mutations (Charlesworth et al. 1993) and the reduction of adaptive potential (erosion of standing variation and reduced fixation of beneficial mutations), the selfing strategy has long been considered as an evolutionary dead-end (Stebbins 1957). The most recent literature has confirmed this prediction, even though the brief duration of selfing lineages often prevent proper testing of the causes of this increased extinction risk (reviewed in Glémin and Galtier 2012; Glémin and Ronfort 2013; Igic and Busch 2013; Wright et al. 2013).

By analyzing the consequences of selfing on a longer timescale, we found evidence of higher accumulation of deleterious mutations in selfing than in outcrossing species, both at the divergence and polymorphism scale. We also found evidence of lack of adaptive response (null or even negative α) and adaptive potential (low polymorphism) in selfing species. *Galba truncatula* and *G. cubensis* thus seem to suffer from the deleterious consequences of selfing, and exhibit all predicted features that should drive them to extinction. Paradoxically, as selfing is likely not of recent origin in the *Galba* clade (see above), this raises the question of the persistence of these species, in apparent contradiction with the prediction of the dead-end hypothesis.

The role of genetic factors, especially the genetic load, in population extinction has been much debated. Under hard-selection, the reduction in fitness due to the accumulation of deleterious alleles can lead populations to extinction when population size decreases below a critical threshold (mutational meltdown; Lynch et al. 1993; Awad et al. 2014). However, under soft-selection, populations could cope with high genetic loads (e.g., Lesecque et al. 2012; Charlesworth 2013, for formal analyses), and put into an ecological context, the load may have no direct consequences on population abundance or persistence (Agrawal and Whitlock 2012). If competition for resources is mainly intraspecific, species could cope with a deteriorated genome: More loaded individuals would be eliminated but resources would be still available for less loaded individuals of the same species. However, under interspecific competition, the least loaded species could exclude the most loaded one (Agrawal and Whitlock 2012). It is possible that *G. truncatula* has been escaping the consequences of the genetic load so far by avoiding interspecific competition. In fact, it is often found in unstable and fluctuating water environments (Trouvé et al. 2005), where interspecific competition is likely to be low. The peculiar ecology of this species can thus explain both the initial evolution of selfing through reproductive assurance and its maintenance because of low competitive pressure. This suggests that ecological context should be taken into account to modulate the dead-end hypothesis. More generally, our results highlight the difficulty of inferring the possible causes of extinction through the analyses of extant species. When selfing species are of recent origin and doomed to rapid extinction, too weak genomic signatures are typically left, while if sufficient time has elapsed to leave detectable genomic signatures, this means that the considered species have escaped extinction until now, and might be exceptions to the dead-end paradox. New framework still needs to be developed to tackle this conundrum.

Materials and Methods

Studied Species and Sample Collection

We focus here on hermaphroditic freshwater snails (Gastropoda: Pulmonata: Hygrophila), an extensively studied animal group because some species are vectors for human and livestock parasites (Brown 1994; Correa et al. 2010). In this group the mating system is well documented, with predominant selfing having independently evolved several times (Escobar et al. 2011). Estimates of selfing rate, inbreeding depression, and waiting time to selfing are available for many species, based on genetic and reproductive biology data. Among the predominantly outcrossing species, we choose *Physa acuta* and *Physa gyrina* (as outgroup), with estimated outcrossing rates greater than 90% (Henry et al. 2005; Escobar et al. 2011) and greater than 70% (Escobar et al. 2011), respectively. Among predominantly selfing species we selected *Galba truncatula*, with estimated outcrossing rates less than 2% (Trouvé et al. 2003; Meunier et al. 2004), and *Galba cubensis* (outgroup) that shows extremely low heterozygosity and polymorphism at 16 microsatellite markers (Lounnas M, Vazquez AA, Hurtrez-Boussès S, unpublished data) and presents morphologic and behavior characteristics of selfing species (Correa et al. 2011). Each species pair is included in a well-supported clade (Wethington and Lydeard 2007; Correa et al. 2010; Dayrat et al. 2011), whose species share similar mating systems according to current knowledge (fig. 1, and references therein). The focal species *P. acuta* and *G. truncatula* share several life-history traits (e.g., adult and egg size), the cosmopolitan distribution (both have presumably spread over all continents from a North American source; Dillon et al. 2002; Correa et al. 2010; Bousset et al. 2014), and the habitat (permanent and ephemeral freshwater environments). Outgroups *P. gyrina* and *G. cubensis* have a narrower distribution, spanning North America and the Neotropics, respectively (Dillon 2000; Correa et al. 2011). Ten individuals of *G. truncatula*, two of *G. cubensis*, nine of *P. acuta*, and two of *P. gyrina* were collected in 2010 and 2011. One individual was sampled per population. The geographical distribution of populations was chosen in order to sample the whole natural distribution range of each species (supplementary table S5, Supplementary Material online). The two *Physa* species were included in a recent analysis of polymorphism levels across outcrossing animals (Romiguier et al. 2014). The *Galba* samples are newly analyzed here.

RNA Extraction and Sequencing

For each individual, total RNA was extracted after removing the shell and digestive tracts to avoid environmental contaminations using standard protocols as described in Gayral et al. (2011), and a nonnormalized cDNA library was prepared from 5 µg RNA. The libraries were sequenced on a Genome Analyzer II or Hiseq 2000 (Illumina, Inc.) to produce 100-bp single-end fragments (Illumina reads). In addition, for one individual of *G. truncatula*, also a normalized random-primed cDNA library was prepared and sequenced for half a run using a 454 Genome Sequencer FLX Titanium Instrument (Roche Diagnostics). Reads were trimmed of low-quality terminal portions using the SeqClean program (http://compbio.dfci.harvard.edu/tgi/, last accessed May 26, 2015).

Transcriptome Assembly, Read Mapping, and Coding Sequence Prediction

De novo transcriptome assembly based on the 454 (one individual in *G.truncatula*) and Illumina reads was performed following strategies B and D in Cahais et al. (2012), using a combination of the programs Abyss and Cap3. Reads were mapped to predicted cDNAs (contigs) using the BWA program, setting the mismatch penalty to 10. Contigs covered at 2.5X (X = diploid sample size) or less across all individuals were discarded. Open-reading frames (ORFs) were predicted using the program transcripts_to_best_scoring_ORFs.pl, which is part of the Trinity package. Contigs carrying no ORF longer than 200 bp were discarded.

Genotype and SNP Calling

At each position of each ORF and each individual, diploid genotypes were called according to the method described by Tsagkogeorga et al. (2012, model M1) and improved by Gayral et al. (2013), implemented in the "reads2snps" program. This method first estimates the sequencing error rate in the maximum-likelihood framework, calculates the posterior probability of each possible genotype, and retains genotypes supported with probability higher than a given threshold (0.95 here)—otherwise missing data are called. We required a minimum coverage of 10X per position and per individual to call a genotype. Then, SNPs were filtered for possible hidden paralogs (duplicated genes) using a likelihood ratio test based on explicit modeling of paralogy ("paraclean" option of the reads2snps software; Gayral et al. 2013). For both the genotype calling and the paralogy filtering procedure, the departure from Hardy–Weinberg expectation was taken into account setting a value of expected heterozygosity, F (extension of reads2snps in Nabholz et al. 2014). First, genotype and SNPs were called assuming panmixia ($F = 0$), and F was estimated on the retained SNPs to confirm selfing and outcrossing status of focal species. As we have species-wide samples, F is equivalent to a F_{IT} and mainly corresponds to F_{IS} for selfing species and to F_{ST} for outcrossing ones. As the assumed expected heterozygosity can affect genotype calling and paralog filtering, reads2snps was run a second time for *G. truncatula* and *G. cubensis* using the F estimated after the first step. Because F was much lower for *P. acuta* we kept the initial genotype calling and filtering procedure in this species.

To verify that ORFs corresponded to known proteins, each retained ORF of each focal species was translated to protein and compared with the nonredundant NR-database using BLASTP, following Romiguier et al. (2014). The first significant hit was recorded (e value < 0.001).

Orthologous Genes Identification and Annotation

Orthologous pairs of ORFs, hereafter called genes, from the focal and the outgroup species were identified using

reciprocal best hits on BLASTn results, a hit being considered as valid when alignment length was above 130 bp, sequence similarity above 80%, and *e* value below e^{-50} (Gayral et al. 2013). Outgroup sequences were added to within-focal species alignments using a profile-alignment version of MACSE (Ranwez et al. 2011), a program dedicated to the alignment of coding sequences and the detection of frameshifts. Genes were only retained if no frameshift was identified by MACSE, and if the predicted ORF in the focal species was longer than 100 codons. To retrieve orthologous genes between the two focal species *P. acuta* and *G. truncatula*, reciprocal best BLAST hit procedure was repeated on the final set of retained genes for each species. Gene annotation was obtained with BLAST2Go version 2.7.1 (Conesa et al. 2005) using default parameters.

Polymorphism and Divergence Statistics

For population genomic statistics, we further filtered the data sets. Positions at which a genotype could be called in less than five individuals for focal species and in less than two individuals for outgroups were discarded. Population statistics were calculated using custom programs that rely on the Bio++ libraries (Guéguen et al. 2013). For each gene, the following statistics were calculated: Per-site synonymous (π_S) and nonsynonymous (π_N) diversity in focal species, heterozygote deficiency (F), number of synonymous (p_S) and nonsynonymous (p_N) segregating sites, number of synonymous (d_S) and nonsynonymous (d_N) fixed differences between focal and outgroup species. These statistics were computed from complete, biallelic sites only—that is, sites showing no missing data after alignment cleaning, and no more than two distinct states. For each species, statistics were averaged across genes weighting by the number of complete sites per gene, thus giving equal weight to every SNP. For p_N/p_S and d_N/d_S, we first computed the averages of p_N, p_S, d_N, and d_S and subsequently the ratios of averages. Confidence intervals were obtained by 10,000 bootstraps over genes. For the selfing species, all statistics were calculated on n/2 alleles, by randomly drawing one haploid sequence per gene and individual. Furthermore, to verify that the observed differences between species are not due to functionally different sets of genes, we recalculated all polymorphism statistics for a subset of orthologous genes between the two focal species and all divergence statistics for a subset of orthologous genes between all four species.

Inference of Demographic History

To test whether recent demographic history could explain the differences in polymorphism patterns between the two species and to help the interpretation of molecular patterns, we inferred the demographic parameters of two simple nested demographic scenarios: 1) A null model with constant population size, and 2) a change from population size N_0 to N_1, T_1 generations in the past. Using outgroups to polarize SNPs, we built the unfolded synonymous SFS, (i.e., the distribution of derived allele counts across SNPs) using custom C++ programs kindly provided by Benoît Nabholz. To cope with variable sample sizes across SNPs, a hypergeometric projection of the observed SFS into a subsample of 12 (for the outcrossing *P. acuta*) and 8 (for the selfing *G. truncatula*) sequences was applied (Hernandez et al. 2007). SNPs sampled in less than 12 (six diploid individuals) or 8 (eight haploid individuals) sequences in *P. acuta* and *G. truncatula*, respectively, were disregarded as far as SFS was concerned. We used the unfolded spectra to fit the models with the diffusion approximation method implemented in the *dadi* software (Gutenkunst et al. 2009). To take possible polarization errors into account, we added an error rate as an extra parameter that was jointly fitted in the model (see *dadi*'s code in supplementary material S2, Supplementary Material online).

Inference of Selection

The global strength of purifying selection was evaluated through the π_N/π_S ratio. Because expression level is usually found to affect the intensity of purifying selection (Drummond et al. 2005; Park et al. 2013; Nabholz et al. 2014), we first controlled for the effect of expression level on π_N/π_S. To avoid the loss of information due to ORFs with null polymorphism, we grouped genes into 100 classes of expression containing a similar number of SNPs and we computed the mean π_N/π_S for each class. We used the average coverage per gene as a proxy for the expression level. Average coverage per gene was calculated by summing the length of matching reads and dividing the result by the cDNA length and the total number of reads sequenced. To test for a species difference in π_N/π_S controlling for expression level, we then performed an ANCOVA including species, expression level, and their interaction. π_N/π_S and expression level were log transformed. For comparative purposes, analyses were repeated for orthologous genes only (not grouped in classes of expression, so only those with π_N and $\pi_S > 0$ were included).

Using all SNPs, even those that could not be polarized, we computed the folded spectra (i.e., the distribution of minor allele count across SNPs) for a sample of 12 (*P. acuta*) and 8 (*G. truncatula*) sequences as explained above (supplementary fig. S6, Supplementary Material online). For each species pair, the synonymous and nonsynonymous folded spectra were used with the "second" Eyre-Walker and Keightley (2009) method to estimate the genome-wide DFE of nonsynonymous mutations, the proportion of adaptive substitution (α), and the rate of adaptive (ω_a) and nonadaptive (ω_d) nonsynonymous substitutions—with $\omega_a = \alpha\ d_N/d_S$ and $\omega_d = (1-\alpha)d_N/d_S$. The method accounts for the effect of demography, or other sources of biases in allele frequencies, by comparing the observed synonymous site-frequency spectrum with the expected one under neutrality (Eyre-Walker et al. 2006). Selection on deleterious mutations was described by a gamma distribution with mean $S = 4N_e s$ and shape parameter β.

Tests of selection were also performed for each gene using two approaches, the MK test (McDonald and Kreitman 1991) and the statistic DoS (Stoletzki and Eyre-Walker 2011). As the MK test quantifies the degree of departure from neutrality with the odds ratio $(p_N/p_S)/(d_N/d_S)$, it cannot be performed

when p_S or d_N are 0. Additionally, the odds ratio tends to be biased and to have a large variance, especially when the number of observations is low (Stoletzki and Eyre-Walker 2011). The statistic $DoS = d_N/(d_N + d_S) - p_N/(p_N + p_S)$ overcomes these limitations (Stoletzki and Eyre-Walker 2011). A positive DoS indicates evidence of adaptive evolution, DoS = 0 indicates neutral evolution, and negative values indicate evidence of purifying selection. An FDR (Storey and Tibshirani 2003) was applied to correct for multiple tests in both analyses ("qvalue" package under R environment).

Given that selective pressures are expected to be relaxed on certain traits in selfing species, some specific genes should show stronger evidence of relaxed selection, particularly genes involved in reproduction (e.g., mating, resource allocation or sexual conflicts; Glémin and Galtier 2012, and references therein). Moreover, genes involved in stress response are more susceptible to be under positive selection (Clark et al. 2007; Kosiol et al. 2008) and are thus a good target to test for the effect of relaxed positive selection in selfers. To investigate if selective pressures differ among genes with different functions, we estimated the average rate of adaptive substitutions separately in three classes of genes: "Reproduction," "response," and "other/unknown." First, we selected genes that were annotated with at least one term of level 2 or that had at least a parent term of level 2 within the domain "Biological process" of the Gene Ontology (GO) term classification. All genes not responding to those criteria were discarded for this analysis. Second, we assigned genes associated with the "reproduction" term (GO:0000003) to the "reproduction" class. Among the remaining genes, we assigned genes associated with any "response" term (GO:0006950, GO:0009605, GO:0009607, GO:0009628, GO:0009719) to the "response" class. Genes not assigned to either the "reproduction or the "response" classes were assigned to the "other/unknown" class. We used a multilocus extension of the MK test implemented in the software MKtest v2.0 (Welch 2006; Obbard et al. 2009). We allowed α (the proportion of adaptive substitutions) and f (the proportion of nonsynonymous mutations that are neutral) to vary among classes, assuming the other parameters of the model to be the same for all classes.

Inference of Selection on Codon Usage and gBGC

We also estimated the intensity of selection on codon usage and gBGC. We first determined codon preferences following Duret and Mouchiroud (1999). The relative synonymous codon usage (RSCU) was calculated for each codon in each gene. The ΔRSCU difference between the most expressed (top 12.5th percentile) and the less expressed (bottom 12.5th percentile) genes was tested for significance with a Mann–Whitney test ($P < 0.05$). Significantly positive ΔRSCU determined preferred codons, significantly negative ΔRSCU determined unpreferred codons. The observed FoP and GC content across all genes were calculated as the proportion of preferred codons over the total number of codons and the proportion of G/C bases over the total number of bases, respectively.

Then, we used unfolded SFS to quantify the strength of selection on codon usage or of gBGC, by estimating the population scaled coefficients $S = 4N_e s$ (where s is the intensity of selection; Kimura 1983) and $B = 4N_e b$ (where b is the intensity of the conversion bias; Nagylaki 1983). Selection on codon usage is expected to increase (respectively decrease) the frequency of unpreferred (U) to preferred (P) mutations (respectively, P to U). Similarly, gBGC is expected to increase (respectively, decrease) the frequency of AT (W for weak) to GC (S for strong) mutations (respectively, S to W). We used the framework of Muyle et al. (2011) to fit a population genetics model on three kinds of spectra: U→P, P→U and neutral for selection on codon usage (U→U and P→P), or W→S, S→W and neutral for gBGC (W→W and S→S). This method takes demographic effects into account as in Eyre-Walker et al. (2006). Because misinference of the ancestral state in the unfolded spectra can give spurious signature of selection or gBGC (Hernandez et al. 2007), we used an extension of this method that takes polarization error into account (Glémin et al. 2014). The method also allows estimating the mutation bias toward AT or unpreferred codons, noted λ here. As for the synonymous and nonsynonymous SFS, we resampled polymorphic sites according to a hypergeometric distribution to obtain a sample of 12 (*P. acuta*) and 8 (*G. truncatula*) sequences. ΔRSCU and frequency spectra were computed using custom PERL scripts kindly provided by Yves Clement. Selection on codon usage and gBGC can be confounded if preferred codons are most often GC-ending. However, selfing has the same consequences on the two processes and we predict the same pattern whatever the underlying process is. We thus did not seek to distinguish the two potential forces.

The obtained estimates of λ, B, and S allowed us to calculate the equilibrium GC content as $GC^* = 1/(1 + \lambda e^{-B})$ and equilibrium FoPs as $FOP^* = 1/(1 + \lambda e^{-S})$ (Bulmer 1991).

Data Accessibility

Data sets are freely available from the Sequence Read Archive (SRA) database (http://http://www.ncbi.nlm.nih.gov/sra) under Study accession SRP056415 and from the Data sets section of the PopPhyl website (http://kimura.univ-montp2.fr/PopPhyl), in which alignment of coding sequences is provided as fasta files.

Supplementary Material

Supplementary materials S1–S3, tables S1–S5, and figures S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the French Centre National de la Recherche Scientifique, by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03), and by European Research Council advanced grant 232971 (PopPhyl). This publication is the contribution ISEM 2015-076 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554). The authors thank Benoît Nabholz and Yves

Clement for providing us scripts for data analyses, Jean-Pierre Pointier for sampling and suggestions on figure 1, and Jonathan Romiguier for BLASTP analysis. They acknowledge D. Rondeleau, J. Dos Santos, Y. Caron, D. Bilton, R. Slapnik, and M. Kalbe for providing samples. They also thank the associate editor and three anonymous reviewers for their helpful comments on the manuscript.

References

Agrawal AF, Whitlock MC. 2012. Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annu Rev Ecol Evol Syst.* 43:115–135.

Arunkumar R, Ness RW, Wright SI, Barrett SCH. 2015. The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics* 199:817-839.

Awad DA, Gallina S, Bonamy C, Billiard S. 2014. The interaction between selection, demography and selfing and how it affects population viability. *PLoS One* 9:e86125.

Awdziejczyk L, Jaeckle W. 2012. Effects of maternal investment on offspring viability in *Physa acuta*. John Wesley Powell Student Research Conference. Illinois Wesleyan University, April 2012. Available from: http://digitalcommons.iwu.edu/cgi/viewcontent.cgi?article=2847&context=jwprc.

Baker HG. 1955. Self-compatibility and establishment after "long-distance" dispersal. *Evolution* 9:347–349.

Baker HG. 1967. Support for Baker's law-as a rule. *Evolution* 21:853–856.

Bargues MD, Artigas P, Khoubbane M, Flores R, Glöer P, Rojas-García R, Ashrafi K, Falkner G, Mas-Coma S. 2011. *Lymnaea schirazensis*, an overlooked snail distorting fascioliasis data: genotype, phenotype, ecology, worldwide spread, susceptibility, applicability. *PLoS One* 6(9):e24567.

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH. 2006. The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. *Mol Biol Evol.* 23:1741–1750.

Bousset L, Pointier J-P, David P, Jarne P. 2014. Neither variation loss, nor change in selfing rate is associated with the worldwide invasion of *Physa acuta* from its native North America. *Biol Invasions.* 16:1769–1783.

Brown DS. 1994. Freshwater snails of Africa and their medical importance. London: Taylor and Francis Ltd.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129:897–907.

Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, Chiari Y, Belkhir K, Ranwez V, Galtier N. 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour.* 12:834–845.

Charlesworth B. 2013. Why we are not dead one hundred times over. *Evolution* 67:3354–3361.

Charlesworth D, Morgan MT, Charlesworth B. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. *Genet Res.* 61:39–56.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289-1303.

Charlesworth D, Vekemans X. 2005. How and when did *Arabidopsis thaliana* become highly self-fertilising. *BioEssays* 27:472–476.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Curr Opin Genet Dev.* 11:685–690.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.

Coleman AW, Vacquier VD. 2014. Exploring the phylogenetic utility of ITS sequences for animals: a test case for Abalone (*Haliotis*). *J Mol Evol.* 54:246–257.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.

Correa AC, Escobar JS, Durand P, Renaud F, David P, Jarne P, Pointier J-P, Hurtrez-Boussès S. 2010. Bridging gaps in the molecular phylogeny of the Lymnaeidae (Gastropoda: Pulmonata), vectors of fascioliasis. *BMC Evol Biol.* 10:381.

Correa AC, Escobar JS, Noya O, Velásquez LE, González-Ramírez C, Hurtrez-Boussès S, Pointier J-P. 2011. Morphological and molecular characterization of Neotropic Lymnaeidae (Gastropoda: Lymnaeoidea), vectors of fasciolosis. *Infect Genet Evol.* 11:1978–1988.

Cutter AD, Wasmuth JD, Washington NL. 2008. Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. *Genetics* 178:2093–2104.

Darwin C. 1876. The effects of cross and self-fertilisation in the vegetable kingdom. London: John Murray.

Dayrat B, Conrad M, Balayan S, White TR, Albrecht C, Golding R, Gomes SR, Harasewych MG, de Frias Martins AM. 2011. Phylogenetic relationships and evolution of pulmonate gastropods (Mollusca): New insights from increased taxon sampling. *Mol Phylogenet Evol.* 59:425–437.

Dillon RT. 2000. The ecology of freshwater molluscs. Cambridge: Cambridge University Press.

Dillon RT, Wethington AR, Rhett JM, Smith TP. 2002. Populations of the European freshwater pulmonate *Physa acuta* are not reproductively isolated from American *Physa heterostropha* or *Physa integra*. *Invert Biol.* 121:226–234.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* 102:14338–14343.

Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 4:e1000071.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A.* 96:4482–4487.

Escobar JS, Auld JR, Correa AC, Alonso JM, Bony YK, Coutellec M-A, Koene JM, Pointier J-P, Jarne P, David P. 2011. Patterns of mating-system evolution in hermaphroditic animals: correlations among selfing rate, inbreeding depression, and the timing of reproduction. *Evolution* 65:1233–1253.

Escobar JS, Cenci A, Bolognini J, Haudry A, Laurent S, David J, Glémin S. 2010. An integrative test of the dead-end hypothesis of selfing evolution in Triticeae (Poaceae). *Evolution* 64:2855–2872.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8:610–618.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol.* 26:2097–2108.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173:891–900.

Fisher RA. 1941. Average excess and average effect of a gene substitution. *Ann Eugen.* 11:53–63.

Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* 25:1–5.

Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, et al. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLoS Genet.* 9:e1003457.

Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N. 2011. Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol Ecol Resour.* 11:650–661.

Gioti A, Stajich JE, Johannesson H. 2013. *Neurospora* and the dead-end hypothesis: genomic consequences of selfing in the model genus. *Evolution* 67:3600–3616.

Glémin S. 2003. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* 57:2678–2687.

Glémin S. 2007. Mating systems and the efficacy of selection at the molecular level. *Genetics* 177:905–916.
Glémin S. 2010. Surprising fitness consequences of GC-biased gene conversion: I. mutation load and inbreeding depression. *Genetics* 185:939–959.
Glémin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L. 2014. Quantification of GC-biased gene conversion in the human genome. bioRxivdoi: http://dx.doi.org/10.1101/010173.
Glémin S, Galtier N. 2012. Genome evolution in outcrossing versus selfing versus asexual species. In: Anisimova M, editor. Evolutionary genomics. Methods in molecular biology. Humana Press. p. 311–335. Available from: http://link.springer.com/protocol/10.1007/978-1-61779-582-4_11.
Glémin S, Muyle A. 2014. Mating systems and selection efficacy: a test using chloroplastic sequence data in Angiosperms. *J Evol Biol.* 27:1386–1399.
Glémin S, Ronfort J. 2013. Adaptation and maladaptation in selfing and outcrossing species: new mutations versus standing variation. *Evolution* 67:225-240.
Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D, Pouyet F, Cahais V, et al. 2013. Bio++: efficient extensible libraries and tools for computational molecular evolution. *Mol Biol Evol.* 30:1745–1750.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695.
Haudry A, Cenci A, Guilhaumon C, Paux E, Poirier S, Santoni S, David J, Glémin S. 2008. Mating system and recombination affect molecular evolution in four Triticeae species. *Genet Res.* 90:97–109.
Hazzouri KM, Escobar JS, Ness RW, Killian Newman L, Randle AM, Kalisz S, Wright SI. 2013. Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection, and evolution of biased gene conversion with an ongoing mating system shift. *Evolution* 67:1263–1278.
Henry P-Y, Bousset L, Sourrouille P, Jarne P. 2005. Partial selfing, ecological disturbance and reproductive assurance in an invasive freshwater snail. *Heredity* 95:428–436.
Heppleston PB. 1972. Life history and population fluctuations of *Lymnaea truncatula* (Mull), the snail vector of fascioliasis. *J Appl Ecol.* 9:235–248.
Hernandez RD, Williamson SH, Zhu L, Bustamante CD. 2007. Context-dependent mutation rates may cause spurious signatures of a fixation bias favoring higher GC-content in humans. *Mol Biol Evol.* 24:2196–2202.
Igic B, Busch JW. 2013. Is self-fertilization an evolutionary dead end? *New Phytol* 198:386–397.
Ingvarsson P. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* 56:2368–2373.
Jarne P. 1995. Mating system, bottlenecks and genetic polymorphism in hermaphroditic animals. *Genet Res.* 65:193–207.
Jarne P, Auld JR. 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution* 60:1816–1824.
Jarne P, Charlesworth D. 1993. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annu Rev Ecol Syst.* 24:441–466.
Kamran-Disfani A, Agrawal AF. 2014. Selfing, adaptation and background selection in finite populations. *J Evol Biol.* 27:1360–1371.
Kawabe A, Fujimoto R, Charlesworth D. 2007. High diversity due to balancing selection in the promoter region of the Medea gene in *Arabidopsis lyrata*. *Curr Biol.* 17:1885–1889.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4:e1000144.
Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. *Trends Ecol Evol.* 29:33–41.
Lesecque Y, Keightley PD, Eyre-Walker A. 2012. A resolution of the mutation load paradox in humans. *Genetics* 191:1321–1330.
Li X-Y, Dong X-Y, Bai X, Liu L, Wang J-J. 2014. The embryonic and postembryonic developmental toxicity of imidazolium-based ionic liquids on *Physa acuta*. *Environ Toxicol.* 29:697–704.
Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates Inc.
Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The mutational meltdown in asexual populations. *J Hered.* 84:339–344.
Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19:330–338.
Marais G, Charlesworth B, Wright SI. 2004. Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biol.* 5:R45.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
Meunier C, Hurtrez-Boussès S, Jabbour-Zahab R, Durand P, Rondelaud D, Renaud F. 2004. Field and experimental evidence of preferential selfing in the freshwater mollusc *Lymnaea truncatula* (Gastropoda, Pulmonata). *Heredity* 92:316–322.
Meunier C, Tirard C, Hurtrez-Boussès S, Durand P, Bargues MD, Mas-Coma S, Pointier JP, Jourdane J, Renaud F. 2001. Lack of molluscan host diversity and the transmission of an emerging parasitic disease in Bolivia. *Mol Ecol.* 10:1333–1340.
Muyle A, Serres-Giardi L, Ressayre A, Escobar J, Glémin S. 2011. GC-biased gene conversion and selection affect GC content in the *Oryza* genus (rice). *Mol Biol Evol.* 28:2695–2706.
Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, Ghesquière A, Santoni S, David J, Glémin S. 2014. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol Ecol.* 23:2210–2227.
Nagylaki T. 1983. Evolution of a large population under gene conversion. *Proc Natl Acad Sci U S A.* 80:5941–5945.
Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929.
Obbard DJ, Welch JJ, Kim K-W, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5:e1000698.
Park C, Chen X, Yang J-R, Zhang J. 2013. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A.* 110:E678–E686.
Pollak E. 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117:353–360.
Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. 2011. Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biol Evol.* 3:868–880.
Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.
Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, Webster MT. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. *Philos Trans R Soc B Biol Sci.* 365:2571–2580.
Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.
Roze D, Rousset F. 2004. Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics* 167:1001–1015.
Schoen DJ, Brown AH. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc Natl Acad Sci U S A.* 88:4494–4497.
Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*,

a plant species with a large effective population size. *Mol Biol Evol.* 27:1813–1821.
Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 45:831–835.
Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo J-M, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive Darwinian selection at the imprinted MEDEA locus in plants. *Nature* 448:349–352.
Stebbins GL. 1957. Self fertilization and population variability in the higher plants. *Am Nat.* 91:337–354.
Stebbins GL. 1974. Flowering plants: evolution above the species level. Cambridge (MA): Belknap Press of Harvard University Press.
Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Mol Biol Evol.* 28:63-70.
Storey JD, Tibshirani R. 2003. Statistical significance for genome-wide studies. *Proc Natl Acad Sci U S A.* 100:9440–9445.
Taylor DW. 1988. Aspects of freshwater mollusc ecological biogeography. *Palaeogeogr Palaeoclimatol Palaeoecol.* 62:511–576.
Trouvé S, Degen L, Goudet J. 2005. Ecological components and evolution of selfing in the freshwater snail *Galba truncatula*. *J Evol Biol.* 18:358–370.
Trouvé S, Degen L, Renaud F, Goudet J. 2003. Evolutionary implications of a high selfing rate in the freshwater snail *Lymnaea truncatula*. *Evolution* 57:2303–2314.
Tsagkogeorga G, Cahais V, Galtier N. 2012. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol Evol.* 4:852–861.
Tsitrone A, Jarne P, David P. 2003. Delayed selfing and resource reallocations in relation to mate availability in the freshwater snail *Physa acuta*. *Am Nat.* 162:474–488.
Vogler DW, Kalisz S. 2001. Sex among the flowers: the distribution of plant mating systems. *Evolution* 55:202–204.
Welch JJ. 2006. Estimating the genomewide rate of adaptive protein evolution in *Drosophila*. *Genetics* 173:821–837.
Wethington AR, Dillon RT Jr. 1997. Selfing, outcrossing, and mixed mating in the freshwater snail *Physa heterostropha*: lifetime fitness and inbreeding depression. *Invertebr Biol.* 116:192–199.
Wethington AR, Lydeard C. 2007. A molecular phylogeny of Physidae (Gastropoda: Basommatophora) based on mitochondrial DNA sequences. *J Molluscan Stud.* 73:241–257.
Whitlock MC, Barton NH. 1997. The effective size of a subdivided population. *Genetics* 146:427–441.
Wright SI, Kalisz S, Slotte T. 2013. Evolutionary consequences of self-fertilization in plants. *Proc R Soc B Biol Sci.* 280:20130133.
Wright SI, Lauga B, Charlesworth D. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol Biol Evol.* 19:1407–1420.
Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. *Int J Plant Sci.* 169:105–118.