Mating systems and recombination landscape strongly shape genetic diversity and selection in wheat relatives

Concetta Burgarella[1,2,3], Marie-Fleur Brémaud[2], Gesa Von Hirschheydt[4], Veronique Viader[2], Morgane Ardisson[2], Sylvain Santoni[2], Vincent Ranwez[2], Jacques David[2] and Sylvain Glémin[5,6]**

[1] CNRS, Univ Montpellier, ISEM– UMR 5554, 34095 Montpellier, France.

[2] AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France.

[3] Department of Organismal Biology, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden.

[4] Swiss Federal Research Institute WSL, Zürcherstrasse 111 8903 Birmensdorf, Switzerland.

[5] CNRS, Univ. Rennes, ECOBIO - UMR 6553, F-35000 Rennes, France.

[6] Department of Ecology and Evolution, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden.

***Corresponding authors**
Concetta Burgarella: Department of Organismal Biology, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden. E-mail: concetta.burgarella@ebc.uu.se

Sylvain Glémin: CNRS, Univ. Rennes, ECOBIO - UMR 6553, F-35000 Rennes, France. Department of Ecology and Evolution, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden. E-mail: sylvain.glemin@univ-rennes1.fr

Keywords: self-fertilization, polymorphism, linked selection, fitness effect of mutations, selfing-syndrome

Abstract

How and why genetic diversity varies among species is a long-standing question in evolutionary biology. Life history traits (especially life-span and reproductive mode) have been shown to be main determinants, but the underlying causes are debated. Demographic history and linked selection have been proposed as main mechanisms, but their relative impact is not well understood. Here, we used the *Aegilops*/*Triticum* grass species, a group characterized by contrasted mating systems (from obligate out-crossing to high selfing) and marked recombination rate variation across the genome, to quantify the interaction between demography and linked selection on neutral and adaptive diversity. By analysing phenotypic and transcriptomic data, we found that selfing strongly affects genetic diversity and the efficacy of selection by amplifying the intensity of linked selection genome wide. Differently from expectations, this effect was gradual, with mixed mating species showing intermediate patterns. These results bear implications for the evolution of mating systems and more generally for our understanding of the fundamental drivers of genetic diversity.

Introduction

How and why genetic diversity varies among species is a central and long-standing question in evolutionary biology, dating back from the 60's (Ellegren & Galtier, 2016). According to population genetic theory, neutral genetic diversity scales with the product of the mutation rate, u, and the effective population size, *Ne*, which measures the amount of genetic drift and corresponds to the size of an idealized population where all individuals contribute equally to reproduction. It follows that genetic diversity should positively correlate with species abundance. However, empirical evidence does poorly supports this prediction; most often, genetic diversity does not correlate with proxies of census of population size or abundance.

In contrast, several studies have shown that life history traits (LHTs), especially life-span and reproductive mode, can explain a large part of the observed variation in genetic diversity among species (Chen et al., 2017; Mackintosh et al., 2019; Muyle et al., 2021; Romiguier et al., 2014). This suggests that LHTs may better reflect long-term *Ne* than recent demographic history. The few exceptions where genetic diversity correlates well with abundance or census size correspond to conditions where population sizes have remained large or small for long periods of time, such as shown in a recent study comparing mainland and island bird species(Leroy et al., 2021).

However, it is not clear how LHTs determine long term *Ne.* LHTs could control the demographic response of a species to environmental perturbations (Romiguier et al., 2014). Alternatively, LHTs, in particular reproductive modes, could modulate linked selection, which also affects long-term *Ne* and seems rather pervasive across genomes (Buffalo, 2021; Chen et al., 2020; Corbett-Detig et al., 2015; Mackintosh et al., 2019). Demography and linked selection affect patterns of genetic diversity but also the efficacy of selection and how it varies among species and across genomes, yet in a different way. Demography is predicted to have genome wide effects whereas linked selection effects can vary with genome structure (e.g. recombination landscape, gene density). Therefore, understanding not only the determinants but also the underlying mechanisms affecting genetic diversity is central to analyze genomic patterns of variation and to understand how species respond to selection.

Among LHTs, mating systems deeply affect the genetic and ecological functioning of a species and is predicted to strongly impact both demography and linked selection. Comparing species with contrasted mating systems can thus be a convenient approach to disentangle the role of the different drivers of genetic diversity. Thanks to the ability to produce seeds under limited mate availability, the capacity of autonomous selfing provides reproductive assurance and can be an ecologically successful strategy, allowing colonizing new habitats and increasing species range (Grossenbacher et al., 2015). However, being able to reproduce alone implies a much higher demographic stochasticity due to recurrent bottlenecks and colonization-extinction dynamics (Ingvarsson, 2002). In addition, selfing also has direct negative genetic effects. Non-independent gamete sampling during mating automatically increases genetic drift and reduces genetic mixing, which generates genome-wide genetic linkage disequilibrium and enhances the effects of linked selection (Burgarella & Glémin, 2017).

So far, striking differences in genetic diversity have already been observed

between outcrossing and selfing relatives (e.g., Burgarella et al., 2015; Hazzouri et al., 2013; Slotte et al., 2013), but the underlying causes (linked-selection, demographic instability) have neither been directly identified nor quantified. Interestingly, linked selection is predicted to be especially enhanced at high selfing rates (Agrawal & Hartfield, 2016; Roze, 2016), whereas bottleneck dynamics can occur as soon as autonomous selfing is possible but not necessarily predominant. Under linked selection effects we thus predict a strong difference in genomic patterns between high selfers and other species, whereas under demographic effects we predict the main difference to be between obligate outcrossers (such as self-incompatible species) and others (self-compatible ones). Comparing species with a large range of mating systems occurring within a single genus allows strong demographic and genetic contrast among otherwise similar species.

Here we investigated the question of the determinants and the mechanisms driving patterns of genetic diversity using the *Aegilops*/*Triticum* grass species as a study system. This group of Mediterranean and Western/Central Asian grasses belongs to the Triticeae tribe (Poaceae) and includes wheat and its wild relatives. The *Aegilops*/*Triticum* genus forms a monophyletic group with 13 diploid and about 17 polyploid species that likely diversified ~4-7 millions years ago(Huang et al., 2002; Marcussen et al., 2014) with a complex past history of hybridization (Bernhardt et al., 2020; Glémin, Scornavacca, et al., 2019; Huynh et al., 2019). All species are annual, wind-pollinated, and present a large diversity of mating systems, spanning from obligate out-crossing to highly selfing species(Kilian et al., 2011; van Slageren, 1994) (**Fig. 1**). Triticeae genomes are large, with markedly U-shaped recombination patterns along chromosomes: most recombination is located in the distal parts whereas no or very low recombination occurs in their central part (Brazier & Glémin, 2022b). Marked differences in both mating systems among species and recombination rate within genomes make the group an ideal model to unravel the factors that may affect species genetic diversity. Through a comparative population genomic approach, we assessed the expectation that genetic diversity and selection efficacy decreases with higher selfing rate and lower species range among species, and with lower recombination rate within each species. We then explicitly tested whether the effect of linked selection was stronger in selfing species as predicted by population genetics theory. We found that selfing strongly and gradually affects genetic diversity and the efficacy of selection by amplifying the intensity of linked selection genome wide, which have multiple implications for both the evolution of mating systems and our understanding of the fundamental drivers of genetic diversity.

Results

Mating system strongly varies in Aegilops/Triticum genus

We analyzed phenotypic and transcriptomic diversity in 98 accessions from the 13 diploid *Aegilops*/*Triticum* species and three close outgroup species *Taenatherium caput-medusae*, *Hordeum vulgare* and *Secale vavilovii* (**Fig. 1** and **Table S1**).

For some species the mating system was already well known, including the self-incompatible *Ae. speltoides* and *Ae. mutica* and the highly selfing *Triticum urartu*, *T. monococcum* and *Ae. Tauschii* (Dvořák et al., 1998; Escobar et al., 2010), but for others it was poorly documented (Kilian et al., 2011). We thus characterized the mating system of

each species by quantifying six floral and reproductive traits, including the size of female and male reproductive organs (anthers, stigmas), male investment, spikelet compactness, and the autonomous seed set (Escobar et al., 2010) (**Supplementary Tables S2, S3 and S5**). We considered these traits as indicative of the selfing syndrome, i.e. the specific changes in flower morphology and function that are expected to occur following the evolution of self-fertilization, especially for anemophilous species (Escobar et al., 2010; Sicard & Lenhard, 2011). The autonomous seed set provided a verification that self-incompatible species *Ae. mutica* and *Ae. speltoides* produced almost no seeds under imposed self-fertilization in the greenhouse (bagged spikes), while all the other species were able to produce seeds (**Supplementary Fig. S1**). All the other traits were significantly negatively (anther and stigma size, spikelet compactness) or positively (male investment) correlated with the autonomous seeds set (**Supplementary Fig. S2**), indicating that the selected traits are good indicators of the mating system of each species.

We summarized this reproductive morpho-functional diversity with a multivariate approach, principal component analysis (PCA). The PCA first axis reflected the differences in mating systems within the *Aegilops*/*Triticum* genus (**Fig. 1B** and **Supplementary Fig. S3**). On one extreme of PC1 there were self-incompatible *Ae. mutica* and *Ae. speltoides*, which showed bigger anthers and stigmas, higher male/female investment, lower spikelet compactness and lower (null) autonomous seed set. On the other extreme there were the predominantly selfing *Ae. tauschii*, *T. urartu* and *T. monococcum*, with opposite states of the traits (**Fig. 1A and 1B**). The other *Aegilops* species showed intermediate values of the multi-trait statistic. This result still held when outgroups were included in the analysis (**Supplementary Fig. S4**).

Species morphology measured by PC1 explained well the genome-wide estimate of inbreeding coefficient, *F,* calculated on the whole transcriptome dataset (R^2=0.83, p-value=1.686e-05, **Fig. 1C**), suggesting that our phenotypic data represented a good proxy of the mating system. Thus, in the following analyses, we used PC1 to describe the selfing syndrome, which avoids using genomic data for both characterizing mating systems and their genomic consequences.

These findings are in agreement with previous knowledge (Dvořák et al., 1998; Escobar et al., 2010) and allowed us to characterize the mating system for the species that lacked outcrossing rate estimations. They also showed that the effects of selfing are gradual, with species exhibiting mixed mating strategies described by intermediate *F* values and phenotypic selfing syndrome. Mapping mating systems on the phylogeny suggested that self-incompatibility was likely ancestral and may have broken only once as all species are self-compatible, except the two external ones (*Ae. mutica* and *Ae. speltoides*). However, several breakdowns of self-incompatibility cannot be excluded. For example, high selfing could have evolved four times independently, in the branches leading to *Ae. tauschii*, *Ae. searsii*, *Ae. uniaristata* and *Triticum* species (**Fig. 1A**).

Polymorphism correlate with mating system but not with species geographic range

Genetic diversity was estimated for each species from whole transcriptome sequencing data. Sequences of 48 accessions were generated and *de novo* assembled in this study and were added to the datasets of (Glémin, Scornavacca, et al., 2019) and (Clément et al., 2017) (**Supplementary Table S1**). Between 19,518 and 28,834 coding

sequences were obtained per species. After filtering, genotype calling was performed on a number of contigs varying from 7,083 (*Ae. searsii*) to 21,706 (*T. caput-medusae*) (**Supplementary Table S4**).

Selfing is expected to reduce neutral genetic diversity (Ingvarsson, 2002; Jarne, 1995; Nordborg, 2000; Pollak, 1987; Schoen & Brown, 1991), here estimated as synonymous polymorphism, π_S. Across species, π_S varied more than one order of magnitude, from 0.0011 for the self-compatible *Ae. searsii* to 0.02 for the self-incompatible *Ae. speltoides*. In general, and according to expectations, outcrossing species harbored higher genetic diversity than self-compatible species. Neutral genetic diversity, π_S and PC1 were significantly correlated across species (R^2 = 0.75, p-value = 0.00014), indicating a gradient in which stronger selfing syndrome corresponds to lower genome-wide neutral diversity (**Fig. 2A**). The correlation was still significant after phylogenetic control (R^2 = 0.67, p-value: 0.00111).

Selfing is also expected to reduce the efficacy of selection, thus leading to higher accumulation of segregating slightly deleterious mutations in selfing than in outcrossing species (Glémin, 2007). In agreement with this prediction, the efficacy of purifying selection estimated by the ratio of non-synonymous to synonymous polymorphism π_N/π_S (Kimura, 1983) was lower for selfing species (max π_N/π_S value 0.21 for *T. urartu*) than for outcrossing ones (0.09 for *Ae. speltoides*). Similarly to neutral diversity, the efficacy of selection was also significantly explained by the selfing syndrome (R^2=0.48, p-value=0.0083; with phylogenetic control R^2=0.35, p-value=0.041; **Fig. 2B**). We further verified that both polymorphism statistics, π_S and π_N/π_S, also correlated with *F* estimates (**Supplementary Fig. S5**).

If species range also affects genetic diversity, we expect widespread species to be more polymorphic than species with restricted geographic range. Using occurrence data from the GBIF database, we estimated the geographic range of each species. In contrast to mating systems, species range was not correlated with either π_S (**Fig. 2C)** or π_N/π_S (**Fig. 2D)**. Such a correlation could be masked by the strong effect of selfing, for example *Ae. tauschii* is highly selfing and has by far the largest species range. However, a linear model with the two effects showed that mating system still significantly explained both π_S (p-value=0.00004) and π_N/π_S (p-value=0.0157) whereas species range did not (**Supplementary Fig. S6**). Overall, these results suggest that the mating system is the main driver of genetic diversity in *Aegilops*/*Triticum* species and overwhelms potential effects of current population sizes.

The effect of linked selection depends on the mating system

We tested the hypothesis that selfing increases the effect of linked selection by comparing polymorphism patterns and recombination along chromosomes. Species-specific recombination maps were not available for most species, so we used the recombination map of the outgroup species *Hordeum vulgare* since synteny and recombination patterns at large scale are well conserved, in particular the sharp difference between a large central region almost devoid of recombination and highly recombining chromosome arms (Brazier & Glémin, 2022a; Mayer et al., 2011; but see Parisod & Badaeva, 2020). In addition, the outgroup has (on average) the same phylogentic distance to every *Aegilops*/*Triticum* species, avoiding to bias the analysis toward species closer to the reference. For this analysis, we restricted the dataset to the genes for which we identified an orthologue in the barley genome (see **Supplementary**

methods).

In all species, synonymous polymorphism was strongly correlated with recombination and presented a U-shaped pattern along chromosomes more or less mirroring the recombination pattern (see **Fig. 3A** and **Supplementary Fig. S13**). However, the higher the selfing rate, the flatter the relationship (**Fig. 3B, Supplementary Fig. S7**), suggesting a strong effect of the mating system on the relationship between polymorphism and recombination, which cannot be explained by a mutagenic effect of recombination alone (**Supplementary Fig. S7**).

To quantify this relationship, we fitted a simple phenomenological model using a logistic function (**Fig. 3B** for the rationale). The slope of the curve at the inflexion point measures the rate at which polymorphism increases with recombination rate. It clearly decreased with PC1, indicating a very strong and gradual effect of selfing rate on linked selection (**Fig. 3C bottom**). From the fit of the logistic curve, we can also obtain the limit of π_S as recombination tends towards infinity, which can be interpreted as the maximum π_S that could be reached in the absence of linked selection, π_{max}. The π_{max}/π_S ratio also increased with PC1, showing that polymorphism reduction due to linked selection is stronger when selfing rate is higher (**Fig. 3C top**).

To go further, we fitted a linked selection model similar to Corbett-Detig et al. (Corbett-Detig et al., 2015) and Elyashiv et al. (Elyashiv et al., 2016), adapted to partially selfing species. We included both background selection and selective sweeps, taking gene density along chromosomes into account (see **Supplementary Methods** for details). For nine species, the estimations of π_{max} showed that the π_{max}/π_S ratio increased with the selfing syndrome, although with higher estimates and less clear pattern than with the phenomenological model. However, the model is difficult to properly fit because of the highly heterogeneous recombination landscape including large genomic regions with very low or no genetic diversity and strong linkage, which generates strong correlation between estimates of π_{max} and the intensity of linked selection. The estimates of the four remaining species (the *Sitopsis* section) were unstable and could be aberrant, depending on filtering and fitting conditions (**Supplementary Fig. S8**).

Accumulation of deleterious mutations and reduced adaptation under selfing and low recombination

Another central prediction of the effect of genetic linkage is that selection should be less efficient in genomic regions of low recombination, which can extend genome wide in highly selfing species. In agreement with this expectation, the efficacy of purifying selection at the genome wide level clearly decreases with selfing rate (**Fig.2**). All speciesalso showed a negative relationship between recombination rates and the π_N/π_S ratio (**Supplementary Fig. S7**), indicating that purifying selection is more efficient in highly recombining regions.

To better characterize how selection efficacy varies with mating system and recombination, we estimated the full distribution of fitness effects (DFE) of mutations - i.e. including both deleterious and beneficial mutations - using the *polyDFE* method that leverages information from unfolded synonymous and non-synonymous site frequency spectra(Tataru et al., 2017). The method requires a sufficient number of chromosomes sampled (say >10) so we applied it only to the four species with the largest sample sizes, which correspond to the extreme mating systems used to calibrate the selfing syndrome: the two self-incompatible *Ae. speltoïdes* and *Ae. mutica* and two highly selfing *Ae.*

tauschii and *T. urartu*. In agreement with the π_N/π_S ratio analysis, we found that the two selfers suffered from a higher load than the two outcrossers, with 22% to 25% of mutations not being efficiently selected against (-10 < *Ne s* < 0), versus only 9 to 15% in outcrossers (**Fig. 4**). We also found a strong difference between regions of low and high recombination (higher vs lower than the median) for all species. However, the difference was stronger in outcrossers (more than two-fold) than in selfers (30-60% difference only) (**Fig. 4**). Interestingly, purifying selection appeared as efficient in low-recombination regions of outcrossing genomes than in high-recombination regions of highly selfing genomes (**Fig. 4**).

Another striking result is that we estimated an adaptive substitution rate not different from zero in the two selfers, but a quite high value in the two outcrossers and only in highly recombining regions (**Fig. 4**). We verified that this signature of positive selection was not due to the spurious effect of GC-biased gene conversion (**Supplementary Fig. S9**), which is known to happen in recombining regions of grass genomes(Muyle et al., 2011). Overall, selection appeared to be much less efficient, both on beneficial and against deleterious mutations, in low-recombination regions of outcrossing species, and throughout the genome of highly selfing ones.

Discussion

We compared the patterns of genetic diversity and selection across the genome in thirteen species with contrasted mating systems. We found far less polymorphism and far less selection efficacy in selfing than in outcrossing species, as observed in previous studies (Burgarella et al., 2015; e.g., Chen et al., 2017; Glémin et al., 2006; Hazzouri et al., 2013; Laenen et al., 2018; Slotte et al., 2013). In addition, we showed that these genomic effects depend on the interplay between linked selection and mating systems, and vary gradually with self-fertilization rates. For this, we leveraged a study design tailored to go beyond global patterns and decipher their underlying causes. First, differently from previous general comparisons among plant species (Chen et al., 2017; Glémin et al., 2006), we compared related species with similar life history traits, ecology and genomic features, which allows more direct testing of the effect of the mating system. Second, we investigated all species of a clade covering a large range of mating systems, including intermediate mixed mating species, whereas previous studies addressing sister (or closely related) species mainly focused on extreme outcrossing vs selfing comparisons (Burgarella et al., 2015; Hazzouri et al., 2013; Slotte et al., 2013). Third, by using recombination maps we quantified the effect of linked selection in a comparative way.

Linked selection appears to be a main mechanism shaping levels of diversity in wild wheats, as it can reduce polymorphism at least to five-fold (**Fig. 3C, Supplementary Fig. S7)** in highly selfing species. This is in agreement with but higher than observed by (Corbett-Detig et al., 2015) who found a quantitatively limited effect of linked selection except in selfing species (see the re-analysis of (Coop, 2016)). Another expected consequence of linkage and selfing is a reduction in selection efficacy, both against deleterious and in favor of beneficial mutations. We observed a striking contrast between regions of high and low recombination in self-incompatible species, with twice more weakly selected deleterious mutations and no beneficial mutation expected to fix in

regions of low recombination (**Fig.4, Supplementary Fig. S7**). In highly selfing species, instead, recombination had a weaker effect and only for deleterious mutations (**Fig. 4)**. Interestingly, the regions of high recombination in selfers exhibited a similar amount of weakly deleterious mutations as regions of low recombination in outcrossers, and we detected no signature of adaptive evolution at all in the two highly selfing species. Overall, the cross comparison between mating systems and recombination levels clearly showed that the main quantitative effect of selfing is due to high linkage and linked selection.

In addition to linked selection, recurrent bottlenecks and extinction-recolonization dynamics could also contribute to lower polymorphism and selection efficacy in self-compatible species (Ingvarsson, 2002). Two observations support additional effects of selfing beyond linked selection. First, there was still a significant negative correlation between PC1 and $\pi_{\max}$, suggesting that, even after removing linked selection, mating systems still affected polymorphism (**Supplementary Fig. S10**). However, as noted above, this conclusion must be viewed with caution as a proper model of linked selection remains to be done to estimate $\pi_{\max}$ more accurately. Second, the effect of linked selection is predicted to be non-linear with selfing rate(Agrawal & Hartfield, 2016; Roze, 2016) and effective size to be strongly reduced only for high selfing rates. Here, the main difference in polymorphism levels was not observed between highly selfing and other species, as expected under linked selection models, but between self-incompatible and other species. This is more in line with demographic dynamics that can result from uniparental reproduction allowed by self-compatibility and the ability to self autonomously when mate is limiting but not necessarily high and constant selfing rate.

Our results help better understand the evolution of selfing species. In the short term, selfing is known to recurrently evolve from outcrossing, depending on reproductive assurance and gene transmission advantage balanced by inbreeding depression, which can be partly purged during such transitions. In the long run, selfing lineages tend to diversify less than outcrossing ones and selfing is considered an evolutionary dead-end, likely because of higher probability of extinction (Goldberg & Igić, 2012; Igic et al., 2008; Stebbins, 1957). The very causes of higher extinction rates in selfers remain unclear but increased load and loss of genetic diversity and adaptive potential are possible drivers. However, the pace at which the effects of selfing manifest is still poorly known, although it is likely rapid as in the selfing *Capsella rubella* recently derived from the self-incompatible *C. Grandiflora* (Slotte et al., 2013) or within the species *Arabis alpina* among populations with contrasted mating systems (Laenen et al., 2018).

In the *Aegilops*/*Triticum* genus, several evolution instances towards different degrees of selfing likely occurred in a short evolutionary time period, which manifests by wide variations in floral traits associated with specific genetic diversity patterns. At the short phylogenetically scale we studied, we thus found a clear signature of the joint evolution of morphological traits and population genomic patterns, suggesting that the negative effects of selfing manifests rapidly and, interestingly, gradually. It is tempting to propose that the strong and rapid deleterious effects of selfing we detected will accelerate the extinction of the most selfing lineages. However, so far there is no approach to properly test this hypothesis (Glémin, François, et al., 2019; Wright et al., 2013). Genomic degradation is certainly a witness of the transition towards selfing but we still do not know whether it is the ultimate cause of selfing lineages extinction.

Finally, our results also bear more general implications, beyond selfing species. In line with previous comparative analyses of polymorphism at the genome scale (Chen et al., 2017; Romiguier et al., 2014), we showed that life history traits, here the mating system, are a much stronger determinant of genetic diversity than proxies for census population sizes, here species range. Species range had no effect after globally controlling for the mating system (**Supplementary Fig. S6)** neither after removing specifically the effect of linked selection (no significant correlation between species range and π_{max}). In contrast to previous studies, however, the range of genetic diversity is particularly large despite species being closely related and recently diverged (about 6 MYA), with similar genomes and life-history traits (wind-pollinated annual herbs) except mating system. In these wild wheats, species nucleotide polymorphism varies with a factor 20 (from 0.0011 to 0.022). In contrast, in a butterfly family that diverged around 120 MYA, with four-fold variation in body mass and two-fold variation in chromosome numbers, only a factor ten was observed (from 0.0044 to 0.043)(Mackintosh et al., 2019). Similarly, among 28 species widely covering the seed-plant phylogeny and life forms (from annual herbs to trees), the observed range was only slightly higher than in wild wheats, with a factor 28 (from 0.00064 to 0.018)(Chen et al., 2017). This points to mating systems as a main determinant of variation in genetic diversity among species.

However, variation in genetic diversity is still much narrower than variation in species range, which varies by a factor 500. This is in line with the “Lewontin’s paradox”, the general observation that genetic diversity varies much less across species than census size does (Charlesworth & Jensen, 2022; Lewontin, 1974). Despite its strong effect, linked selection is unlikely to explain the limited range of variation in π_S. Although we faced some difficulties in properly fitting a linked selection model, our results suggest that linked selection can reduce polymorphism at least by a factor five, but probably not much more than one or two orders of magnitude (**Supplementary Fig. S8**). Our results reinforce previous observations that, despite being quite pervasive, linked selection alone is unlikely to explain Lewontin’s paradox (Buffalo, 2021; Coop, 2016; Corbett-Detig et al., 2015). Paradoxically, here, linked selection could on the contrary extend the range of variation because its intensity strongly varies with the mating system. For example, π_{max} is less variable (from 0.0041 to 0.035, factor 8) than π_S, but this should be confirmed with more accurate estimations of π_{max}. In any case, selfing blurs the relationship between census and effective population size. Transitions from outcrossing to selfing are often associated with rapid increase in geographic range and census population size allowed by reproductive assurance and uniparental reproduction (Grossenbacher et al., 2015), which could predict higher effective population size and genetic diversity. However, at the same time, increased linked selection generates the opposite effect. In addition, in selfers, range expansion mainly occurs through recurrent bottlenecks and founder events, which can reduce effective size despite increasing census size. For example, the self-compatible common weed shepherd's purse (*Capsella bursa-pastoris*) originated recently (~150,000 years) through allopolyploidization and colonized the whole Eurasia but exhibits a genetic signature of almost continuous decline of its effective size (Kryvokhyzha et al., 2019).

Overall, our results stress that jointly estimating the impact of complex demography and linked selection would be key to better explain patterns of genetic diversity, but it remains a difficult and challenging task (Johri et al., 2022; J. Li et al.,

2012).

Material and methods

Plant material

We analyzed molecular and morphological data of the 13 wild diploid *Aegilops*/*Triticum* species and 3 outgroup species (*Taeniatherum caput-medusae*, *Hordeum spontaneum* and *Secale strictum*) for a total of 98 accessions. We targeted 7-20 individuals from each of five focal species (self-incompatible *Ae. speltoides* and *Ae. mutica* and predominantly selfing *Ae. tauschii*, *T. urartu* and *T. monococcum)*. We included 1-4 accessions from each of the other *Aegilops*/*Triticum* species and outgroups (**Supplementary Methods** and **Table S1**).

Morphological data

We measured morpho-functional traits describing the reproductive organs and function (Escobar et al., 2010; Friedman & Harder, 2005) (**Supplementary Methods**). For each of 5-7 accessions per *Aegilops*/*Triticum* species and 1-3 accessions per outgroup species, we measured: a mature spike (length, spikelet number, grain number), three spikelets per spike (spikelet length), three flowers from one spikelet (length of palea and lemma), 6 anthers (length and width) and 3 stigmas and ovaries (length of each organ). Measures were manually recorded on millimeter paper or taken on photographs with the software analySIS (Soft Imaging System GmbH 2002; **Supplementary Fig. S11**). Missing data (31%) were imputed with the missMDA package(Josse & Husson, 2016) under the R environment (R Core Team, 2018) **(Supplementary Tables S2, S3** and **S5**).

Additional variables were calculated on imputed measures: mean values of anther and stigma dimensions standardized by the flower length, autonomous seed set, spikelet compactness and male investment (**Supplementary Methods**). We used these additional variables to summarize the selfing syndrome with the first axis of a Principal Component Analysis (PCA) (**Supplementary Tables S3** and **S5**) performed with the ade4 package (55) under the R environment.

Species range

To estimate species range, we retrieved occurrence data from the Global Biodiversity Information Facility (http://www.gbif.org) for each species. We manually cleaned the data set to remove occurrences outside species range, which can be due either to identification errors or recent introductions. Cleaned data were mapped on the world map (focusing on western Eurasia and North Africa) on which we applied a grid with cell size of one decimal degree square (~10,000 km^2). We estimated species range as the number of cells occupied by a species time 10,000 km^2.

Sequencing

Full transcriptome sequences were produced for 48 new accessions and added to the dataset of Glémin et al. (Glémin, Scornavacca, et al., 2019) and Clément et al. (Clément et al., 2017). Full transcriptome extraction, quantification, library preparation and sequencing procedures followed Sarah et al. (Sarah et al., 2017) and Glémin et al.

(Glémin, Scornavacca, et al., 2019). RNAs were extracted and prepared separately for leaves and inflorescence tissues, and mixed subsequently in 20% and 80% proportions, respectively. Several quality checks were performed and finally library fragments of 250-300 bp were sequenced with Illumina paired-end technology on a HiSeq3000 sequencer (GenoToul platform, INRA Toulouse, France http://www.genotoul.fr).

Transcriptome assembly, mapping and genotype calling

Reads cleaning and assembly were performed following Sarah et al. (Sarah et al., 2017). CDS per contig was predicted with *prot4est* (Wasmuth & Blaxter, 2004) and three gene datasets: the output of a Rapsearch (Ye et al., 2011) similarity analysis, *Oryza* matrix model for de-novo-based predictions and the codon usage bias observed in *T. monococcum*. We run Rapsearch in either plant species of Uniprot swissprot (http://www.uniprot.org) or in the Monocotyledon species of greenphyl (http://www.greenphyl.org/cgi-bin/index.cgi). For each species, mapping was done following Sarah et al. (Sarah et al., 2017), except for the use of the bwa (H. Li & Durbin, 2009) option –mem. Reads were mapped on the sequences of the individual with the highest coverage or with the highest number of annotated contigs and only on predicted CDS longer than 250 bp (**Supplementary Table S4)**.

Diploid genotypes were called with reads2snps v. 2.0.64 (Gayral et al., 2013; Tsagkogeorga et al., 2012)(available at https://kimura.univ-montp2.fr/PopPhyl/index.php?section=tools). Genotypes with minimum coverage of 10X per position and individual and probability higher than 0.95 were retained. SNPs were then filtered for possible hidden paralogs (Gayral et al., 2013). First, genotype and SNPs were called assuming panmixia (heterozygote deficiency $F = 0$), and F was estimated on the retained SNPs. 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 were run a second time for each species using the F estimated after the first step. For the outgroup species with sample size n=1 (*T. caput-medusae*) we kept the initial genotype calling and filtering procedure.

Open-reading frames (ORFs) were predicted using the program ORF_extractor.pl (https://kimura.univ-montp2.fr/PopPhyl/index.php?section=tools). Gene length and number of SNPs in the final data set per species are provided in **Supplementary Table S4**.

SNPs derived and ancestral status were estimated for the five focal species using the outgroup *T. caputmedusae*. Orthologous genes were identified using reciprocal best hits on BLASTn results setting *e* value below e-50. Outgroup sequences were added to within-focal species alignments using MACSE v. 1.2 (Ranwez et al., 2011), retaining only genes longer than 100 codons and without frameshifts .

Sequence polymorphism analysis

Polymorphism and divergence statistics were calculated with dNdSpNpS v.3. (https://kimura.univ-montp2.fr/PopPhyl/index.php?section=tools). We discarded positions at which a genotype could be called in less than five individuals for species with sample size n>=5 and in less than n/2 for species with n<5. Subsequently, for genes with more than 10 codons, statistics were calculated on complete (no-missing data), biallelic sites: per-site synonymous (S) and nonsynonymous (N) diversity, heterozygote deficiency (F), number of synonymous (π_S) and nonsynonymous (π_N) segregating sites, number of synonymous (D_S) and nonsynonymous (D_N) fixed differences between focal and outgroup

species. 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 π_N/π_S and dN/dS, we first computed the averages of π_N, π_S, dN, and dS and subsequently the ratios of averages. Confidence intervals were obtained by 10,000 bootstraps over genes. For the focal selfing species *Ae. tauschii*, *T. monococcum* and *T. urartu*, all statistics were calculated on n/2 alleles, by randomly drawing one haploid sequence per gene and individual (**Supplementary Table S4**).

Phenomenological analysis of the effect of linked selection

We regressed π_S against recombination rate, after grouping genes into 50 quantiles of recombination, using a logistic function of the form:

$$\pi_S = \frac{a}{1+b\,e^{-cr}}$$

From this, we estimated i) the slope at the inflexion point, given by $\frac{ac}{4}$, which measures the rate at which polymorphism increases with recombination, so lower slopes correspond to stronger effect of linked selection, and ii) a, which is the maximum π_S that can be reached, assuming free recombination so no effect of linked selection. We fitted this non-linear model with the *nls* function of the *R* software using the inverse of the variance of π_S estimates as weights, that is $\frac{n}{\pi_S(1-\pi_S)}$. We reported mean estimates and confidence intervals obtained with 1000 bootstrapped datasets, after discarding non-convergent or aberrant runs.

Fit of an explicit linked-selection model

To go further, we fitted an explicit linked-selection model, taking both background selection and selective sweeps into account, following Corbett-Detig et al. (Corbett-Detig et al., 2015) and Elyashiv et al. (Elyashiv et al., 2016) but including the effect of partial selfing. The genome was split into genomic windows. For each region where $\pi_S(i)$ has been estimated on n_i positions, we assumed that $n_i\pi_S(i)$ followed a binomial distribution with parameter n_i and p_i given by:

$$p_i = \frac{\pi_{max}}{B_i + S_i}$$

where $B_i = e^{\sum_j u_j \frac{s}{(r_{ij}+s)^2}}$, with s the mean selection coefficient against deleterious mutations, is the term of background selection and $S_i = \sum_j v_j \rho\, e^{-\tau r_{ij}}$, with ρ the rate and τ the mean time of sweeps, is the term of selective sweep. r_{ij} represents the probability of recombination between the focal region i and any other region of the genome, j containing L_j coding positions so that $u_j = u\ L_j$ and where u and v are the rates of deleterious and beneficial mutations, respectively. r_{ij} is obtained from the genetic distance, d_{ij}, using Haldane's mapping function: $r_{ij} = (1 - \exp(-2d_{ij}))/2$. To take partial selfing into account, we rescaled r_{ij} by $r_{ij}(1 - F)$ (see 26). Because estimates of F were not very precise, we also run the model by letting F free and being estimated jointly with the

other parameters. The log likelihood function was optimized with the *optim* function in *R* using the “L-BFGS-B” method. We split the genome into regions of equal genetic size (in cM) based on the Marey map interpolation. We tested three window sizes: 0.1, 0.5, and 1 cM and we discarded regions with less than 1000 bp to avoid too noisy data.

Estimation of the distribution of fitness effect of mutations

In the two self-incompatible species, *Ae. mutica* and *Ae. speltoides*, and in two highly selfing species, *Ae. tauschii* and *T. urartu*, we estimated the distribution of fitness effect of mutations from the unfolded synonymous and non-synonymous site frequency spectra using the *PolyDFE* method (Tataru et al., 2017; Tataru & Bataillon, 2019) with a model averaging approach (as in Muyle et al., 2021). Confidence intervals were obtained by bootstrapping SNPs 1000 times. For the four species, uSFS were polarized using *T. caput-medusae* as an outgroup. We performed three analyses: on the whole set of SNPs or by splitting the dataset into two subsets: SNPs from genes in the high or low recombining regions. GC-biased gene conversion (gBGC), a recombination associated processes mimicking selection in favor of G and C nucleotides, is known to be active in grasses (Clément et al., 2017; Rodgers-Melnick et al., 2016) and could lead to spurious signatures of positive selection in highly recombining regions. To test for the possible effect of gBGC, we also re-run the analyses on three categories of SNPs: AT→GC, GC→AT, and G←→C + A←→T.

References

Agrawal, A. F., & Hartfield, M. (2016). Coalescence with Background and Balancing Selection in Systems with Bi- and Uniparental Reproduction: Contrasting Partial Asexuality and Selfing. *Genetics*, *202*(1), 313–326. https://doi.org/10.1534/genetics.115.181024

Bernhardt, N., Brassac, J., Dong, X., Willing, E.-M., Poskar, C. H., Kilian, B., & Blattner, F. R. (2020). Genome-wide sequence information reveals recurrent hybridization among diploid wheat wild relatives. *The Plant Journal*, *102*(3), 493–506. https://doi.org/10.1111/tpj.14641

Brazier, T., & Glémin, S. (2022a). *Diversity and determinants of recombination landscapes in flowering plants* (p. 2022.03.10.483889). bioRxiv. https://doi.org/10.1101/2022.03.10.483889

Brazier, T., & Glémin, S. (2022b). Diversity and determinants of recombination landscapes in flowering plants. *PLOS Genetics*, *18*(8), e1010141. https://doi.org/10.1371/journal.pgen.1010141

Buffalo, V. (2021). Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. *ELife*, *10*, e67509. https://doi.org/10.7554/eLife.67509

Burgarella, C., Gayral, P., Ballenghien, M., Bernard, A., David, P., Jarne, P., Correa, A., Hurtrez-Boussès, S., Escobar, J., Galtier, N., & Glémin, S. (2015). Molecular Evolution of Freshwater Snails with Contrasting Mating Systems. *Molecular Biology and Evolution*, *32*(9), 2403–2416. https://doi.org/10.1093/molbev/msv121

Burgarella, C., & Glémin, S. (2017). Population Genetics and Genome Evolution of Selfing Species. In John Wiley & Sons Ltd (Ed.), *ELS* (pp. 1–8). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0026804

Charlesworth, B., & Jensen, J. D. (2022). How Can We Resolve Lewontin's Paradox? *Genome Biology and Evolution*, *14*(7), evac096. https://doi.org/10.1093/gbe/evac096

Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution*, *34*(6), 1417–1428. https://doi.org/10.1093/molbev/msx088

Chen, J., Glémin, S., & Lascoux, M. (2020). From Drift to Draft: How Much Do Beneficial Mutations Actually Contribute to Predictions of Ohta's Slightly Deleterious Model of Molecular Evolution? *Genetics*, *214*(4), 1005–1018. https://doi.org/10.1534/genetics.119.302869

Clément, Y., Sarah, G., Holtz, Y., Homa, F., Pointet, S., Contreras, S., Nabholz, B., Sabot, F., Sauné, L., Ardisson, M., Bacilieri, R., Besnard, G., Berger, A., Cardi, C., Bellis, F. D., Fouet, O., Jourda, C., Khadari, B., Lanaud, C., … Glémin, S. (2017). Evolutionary forces affecting synonymous variations in plant genomes. *PLOS Genetics*, *13*(5), e1006799. https://doi.org/10.1371/journal.pgen.1006799

Coop, G. (2016). Does linked selection explain the narrow range of genetic diversity across species? *BioRxiv*, 042598. https://doi.org/10.1101/042598

Corbett-Detig, R. B., Hartl, D. L., & Sackton, T. B. (2015). Natural Selection Constrains Neutral Diversity across A Wide Range of Species. *PLOS Biology*, *13*(4), e1002112. https://doi.org/10.1371/journal.pbio.1002112

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*, 1–20.

Dvořák, J., Luo, M.-C., & Yang, Z.-L. (1998). Restriction Fragment Length Polymorphism and Divergence in the Genomic Regions of High and Low Recombination in Self-Fertilizing and Cross-Fertilizing Aegilops Species. *Genetics*, *148*(1), 423–434. https://doi.org/10.1093/genetics/148.1.423

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews. Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Elyashiv, E., Sattath, S., Hu, T. T., Strutsovsky, A., McVicker, G., Andolfatto, P., Coop, G., & Sella, G. (2016). A Genomic Map of the Effects of Linked Selection in Drosophila. *PLOS Genet*, *12*(8), e1006130. https://doi.org/10.1371/journal.pgen.1006130

Escobar, J. S., 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*(10), 2855–2872. https://doi.org/10.1111/j.1558-5646.2010.01045.x

Friedman, J., & Harder, L. D. (2005). Functional associations of floret and inflorescence traits among grass species. *American Journal of Botany*, *92*(11), 1862–1870. https://doi.org/10.3732/ajb.92.11.1862

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., & Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLoS Genetics*, *9*(4), e1003457. https://doi.org/10.1371/journal.pgen.1003457

Glémin, S. (2007). Mating systems and the efficacy of selection at the molecular level. *Genetics*, *177*(2), 905–916. https://doi.org/10.1534/genetics.107.073601

Glémin, S., Bazin, E., & Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1604), 3011–3019. https://doi.org/10.1098/rspb.2006.3657

Glémin, S., François, C. M., & Galtier, N. (2019). Genome Evolution in Outcrossing vs. Selfing vs. Asexual Species. In M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 331–369). Springer. https://doi.org/10.1007/978-1-4939-9074-0_11

Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., Sarah, G., Santoni, S., David, J., & Ranwez, V. (2019). Pervasive hybridizations in the history of wheat relatives. *Science Advances*, *5*(5), eaav9188. https://doi.org/10.1126/sciadv.aav9188

Goldberg, E. E., & Igić, B. (2012). Tempo and Mode in Plant Breeding System Evolution. *Evolution*, *66*(12), 3701–3709. https://doi.org/10.1111/j.1558-5646.2012.01730.x

Grossenbacher, D., Runquist, R. B., Goldberg, E. E., & Brandvain, Y. (2015). Geographic range size is predicted by plant mating system. *Ecology Letters*, *18*(7), 706–713. https://doi.org/10.1111/ele.12449

Hazzouri, K. M., Escobar, J. S., Ness, R. W., Killian Newman, L., Randle, A. M., Kalisz, S., & Wright, S. I. (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*(5), 1263–1278. https://doi.org/10.1111/evo.12027

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., & Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences*, *99*(12), 8133–8138. https://doi.org/10.1073/pnas.072223799

Huynh, S., Marcussen, T., Felber, F., & Parisod, C. (2019). Hybridization preceded radiation in diploid wheats. *Molecular Phylogenetics and Evolution*, *139*, 106554. https://doi.org/10.1016/j.ympev.2019.106554

Igic, B., Lande, R., & Kohn, J. R. (2008). Loss of Self-Incompatibility and Its Evolutionary Consequences. *International Journal of Plant Sciences*, *169*(1), 93–104. https://doi.org/10.1086/523362

Ingvarsson, Pärk. (2002). A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution*, *56*(12), 2368–2373. https://doi.org/10.1111/j.0014-3820.2002.tb00162.x

Jarne, P. (1995). Mating system, bottlenecks and genetic polymorphism in hermaphroditic animals. *Genetics Research*, *65*(03), 193–207. https://doi.org/10.1017/S0016672300033279

Johri, P., Aquadro, C. F., Beaumont, M., Charlesworth, B., Excoffier, L., Eyre-Walker, A., Keightley, P. D., Lynch, M., McVean, G., Payseur, B. A., Pfeifer, S. P., Stephan, W., & Jensen, J. D. (2022). Recommendations for improving statistical inference in population genomics. *PLOS Biology*, *20*(5), e3001669. https://doi.org/10.1371/journal.pbio.3001669

Josse, J., & Husson, F. (2016). missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software*, *70*(1), Article 1. https://doi.org/10.18637/jss.v070.i01

Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., Hammer, K., & Özkan, H. (2011). Aegilops. In C. Kole (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources: Cereals* (pp. 1–76). Springer. https://doi.org/10.1007/978-3-642-14228-4_1

Kimura, M. (1983). Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution*, *1*(1), 84–93.

Kryvokhyzha, D., Salcedo, A., Eriksson, M. C., Duan, T., Tawari, N., Chen, J., Guerrina, M., Kreiner, J. M., Kent, T. V., Lagercrantz, U., Stinchcombe, J. R., Glémin, S., Wright, S. I., & Lascoux, M. (2019). Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid Capsella bursa-pastoris (Brassicaceae). *PLOS Genetics*, *15*(2), e1007949. https://doi.org/10.1371/journal.pgen.1007949

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., Steige, K. A., Kourmpetis, Y., Odong, T., Drouzas, A. D., Bink, M. C. A. M., Ågren, J., Coupland, G., & Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. *Proceedings of the National Academy of Sciences*, *115*(4), 816–821. https://doi.org/10.1073/pnas.1707492115

Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A. E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J. C., De Swardt, D. H., Blanco, G., Thébaud, C., Milá, B., & Nabholz, B. (2021). Island songbirds as windows into evolution in small populations. *Current Biology*, *31*(6), 1303-1310.e4. https://doi.org/10.1016/j.cub.2020.12.040

Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P., & Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: Where do we stand and where could we go? *Molecular Ecology*, *21*(1), 28–44. https://doi.org/10.1111/j.1365-294X.2011.05308.x

Mackintosh, A., Laetsch, D. R., Hayward, A., Charlesworth, B., Waterfall, M., Vila, R., & Lohse, K. (2019). The determinants of genetic diversity in butterflies. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-11308-4

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., The International Wheat Genome Sequencing Consortium, Jakobsen, K. S., Wulff, B. B. H., Steuernagel, B., Mayer, K. F. X., & Olsen, O.-A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, *345*(6194), 1250092. https://doi.org/10.1126/science.1250092

Mayer, K. F. X., Martis, M., Hedley, P. E., Šimková, H., Liu, H., Morris, J. A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubaláková, M., Suchánková, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., … Stein, N. (2011). Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *The Plant Cell*, *23*(4), 1249–1263. https://doi.org/10.1105/tpc.110.082537

Muyle, A., Martin, H., Zemp, N., Mollion, M., Gallina, S., Tavares, R., Silva, A., Bataillon, T., Widmer, A., Glémin, S., Touzet, P., & Marais, G. A. B. (2021). Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus Silene. *Molecular Biology and Evolution*, *38*(3), 805–818. https://doi.org/10.1093/molbev/msaa229

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). *Molecular Biology and Evolution*, *28*(9), 2695–2706. https://doi.org/10.1093/molbev/msr104

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An Ancestral recombination graph with partial self-fertilization. *Genetics*, *154*(2), 923–929.

Parisod, C., & Badaeva, E. D. (2020). Chromosome restructuring among hybridizing wild wheats. *New Phytologist*, *226*(5), 1263–1273. https://doi.org/10.1111/nph.16415

Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, *117*(2), 353–360.

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS ONE*, *6*(9), e22594. https://doi.org/10.1371/journal.pone.0022594

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., & Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences*, *113*(22), E3177–E3184. https://doi.org/10.1073/pnas.1525244113

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. a.-T., Weinert, L. A., Belkhir, K., Bierne, N., … Galtier, N. (2014). Comparative population genomics in animals uncovers the

determinants of genetic diversity. *Nature*, *515*(7526), 261–263. https://doi.org/10.1038/nature13685
Roze, D. (2016). Background Selection in Partially Selfing Populations. *Genetics*, *203*(2), 937–957. https://doi.org/10.1534/genetics.116.187955
Sarah, G., Homa, F., Pointet, S., Contreras, S., Sabot, F., Nabholz, B., Santoni, S., Sauné, L., Ardisson, M., Chantret, N., Sauvage, C., Tregear, J., Jourda, C., Pot, D., Vigouroux, Y., Chair, H., Scarcelli, N., Billot, C., Yahiaoui, N., … Ruiz, M. (2017). A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives. *Molecular Ecology Resources*, *17*(3), 565–580. https://doi.org/10.1111/1755-0998.12587
Schoen, D. J., & Brown, A. H. (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences*, *88*(10), 4494–4497. https://doi.org/10.1073/pnas.88.10.4494
Sicard, A., & Lenhard, M. (2011). The selfing syndrome: A model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, *107*(9), 1433–1443. https://doi.org/10.1093/aob/mcr023
Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A. E., Escobar, J. S., Newman, L. K., Wang, W., Mandáková, T., Vello, E., Smith, L. M., Henz, S. R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., … Wright, S. I. (2013). The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, *45*(7), 831–835. https://doi.org/10.1038/ng.2669
Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *American Naturalist*, *91*, 337–354. CABDirect2. https://doi.org/10.1086/281999
Tataru, P., & Bataillon, T. (2019). polyDFEv2.0: Testing for invariance of the distribution of fitness effects within and across species. *Bioinformatics*, *35*(16), 2868–2869. https://doi.org/10.1093/bioinformatics/bty1060
Tataru, P., Mollion, M., Glémin, S., & Bataillon, T. (2017). Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. *Genetics*, *207*(3), 1103–1119. https://doi.org/10.1534/genetics.117.300323
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 Biology and Evolution*, *4*(8), 852–861. https://doi.org/10.1093/gbe/evs054
van Slageren, M. W. (1994). Wild wheats: A monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). *Wageningen Agricultural University Papers*, *94–7*.
Wasmuth, J. D., & Blaxter, M. L. (2004). prot4EST: Translating Expressed Sequence Tags from neglected genomes. *BMC Bioinformatics*, *5*(1), 187. https://doi.org/10.1186/1471-2105-5-187
Wright, S. I., Kalisz, S., & Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1760), 20130133. https://doi.org/10.1098/rspb.2013.0133
Ye, Y., Choi, J.-H., & Tang, H. (2011). RAPSearch: A fast protein similarity search tool for short reads. *BMC Bioinformatics*, *12*(1), 159. https://doi.org/10.1186/1471-2105-12-159

Acknowledgements and funding sources

We thank G. Sarah, Y. Holtz and P. Joncour for help with bioinformatic analyses and T. Bataillon, M. Lascoux and D. Schoen for helpful discussions and suggestions on the manuscript. The work was fundend by the French Agence Nationale de la Recherche (ANR) (ANR-11-BSV7-013-03). CB has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 839643.

Data availability

Morphological traits measures underlying Fig.1, Fig. 2, Fig. 3 and Supplementary Figures S1-S4, S8, S10-S11 are provided as Supplementary Data1 and Supplementary Data2. Raw data are deposited at the Sequence Read Archives (SRA) under project PRJNA945064.

Supplementary Information

Extended Materials and Methods

Figures S1 to S13

Supplementary tables:

- Supplementary Table S1. Passport information for the accessions of Triticum/Aegilops species and outgroups used in this study.
- Supplementary Table S2. Dataset 1: Morphofunctional traits linked to reproduction measured in 13 Triticum/Aegilops species and three outgroups.
- Supplementary Table S3. Dataset 2: Morphofunctional traits linked to reproduction measured in 13 Triticum/Aegilops species and three outgroups where missing data were imputed with package missMDA under R environment (see Materials and Methods).
- Supplementary Table S4. Sequencing and genome summary statistics per species (13 Triticum/Aegilops species and three outgroups).
- Supplementary Table S5. List of morphological traits measured and used for the analysis

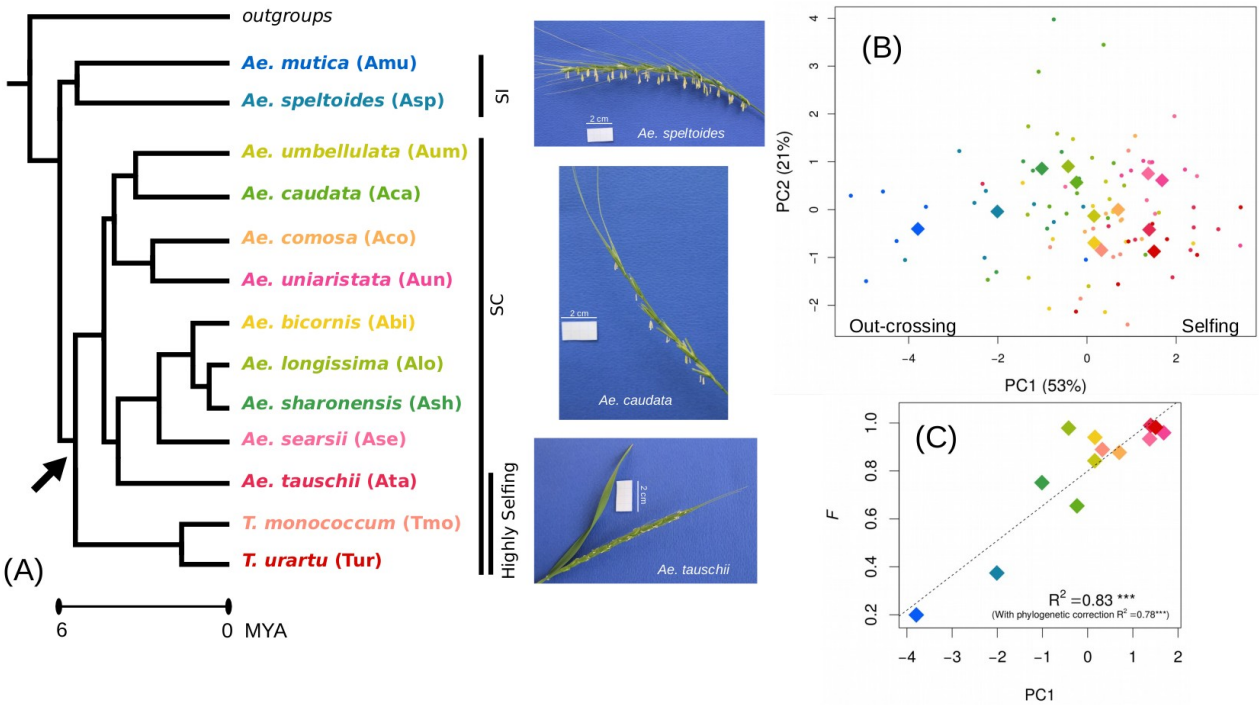

Figure 1. **Mating system diversity in the genus *Aegilops*/*Triticum***. (A) Phylogenetic relationships among the 13 diploid species of the genus, following (Glémin, Scornavacca, et al., 2019). Colors have been assigned to represent a gradient of self fertilization: *Ae. speltoides* and *Ae. mutica* have previously been described as self-incompatible (SI, blue tone colors), *Ae. tauschii* and *Triticum* sp. are known to be prevalently self-fertilizing species (pink-purple colors), while the remaining species are self-compatible (SC, yellow-green colors). The arrow indicates when self-compatibility may have appeared, if we assume that it appeared once according to a parsimony approach. Pictures illustrate the phenotypes associated with different mating systems: SI *Ae. speltoides*, highly selfing *Ae. tauschii* and self-compatible *Ae. caudata*. (B) Principal Component Analysis of morphological measures of reproductive organs. PC1 resumes the morphological selfing syndrome. On one side there are the self-incompatible species, which show bigger anther and stigmas, higher male investment, lower spikelet compactness and lower autonomous seed set. On the other side there are predominately selfing species, with opposite states of the traits. (C) Positive correlation between PC1 of reproductive traits and the inbreeding coefficient *F* estimated on the genomic data shows that species with stronger morphological selfing syndrome have also higher *F* values.

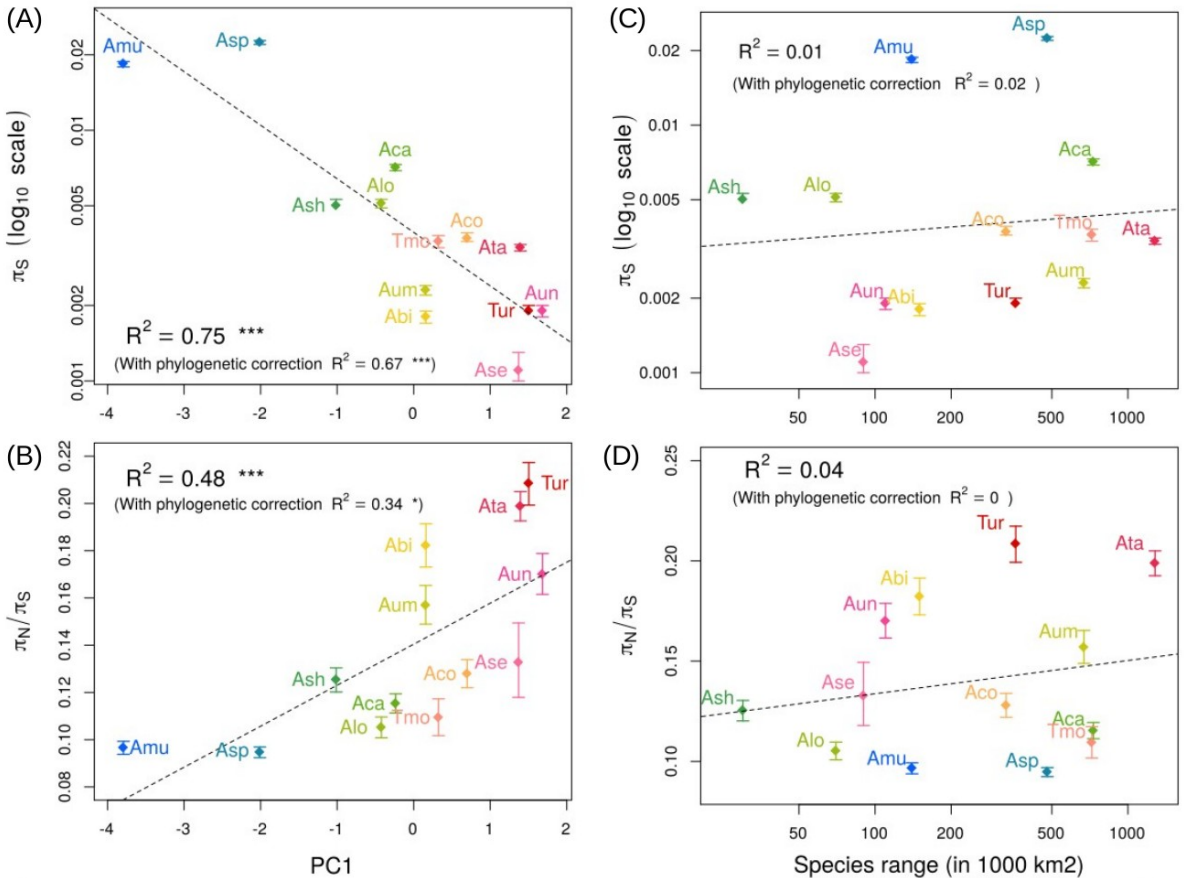

Figure 2. Mating systems traits and global polymorphism patterns. Neutral genetic diversity π_S is negatively correlated with the first axis of a PCA of morphofunctional traits associated with reproduction (A), while the efficacy of selection estimated with π_N/π_S ratio is positively correlated with the same PC1 axis (B). Species geographical range estimated from GBIF occurrence data cannot explain the variation observed neither in π_S (C) nor π_N/π_S (D).

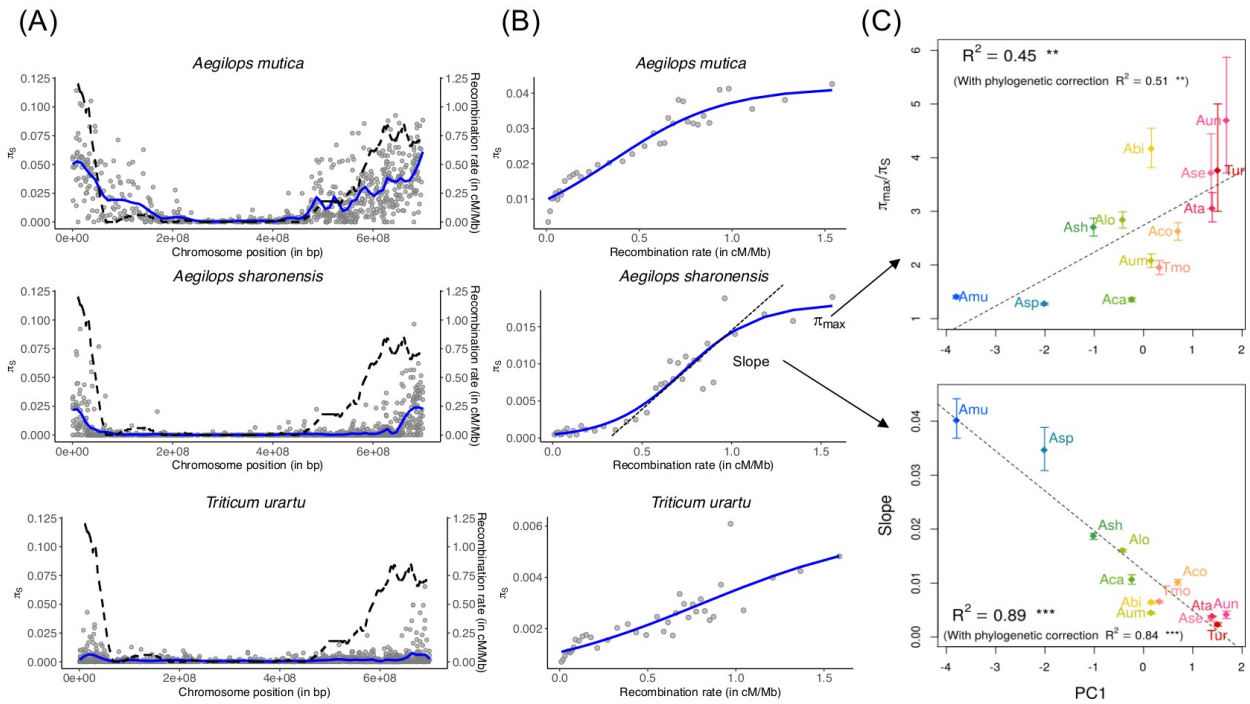

Figure 3. **Recombination and patterns of genetic polymorphism across the genome.** (A) Mean synonymous diversity (π_S) along chromosome 3 for three species with contrasted mating system: self-incompatible (*Ae. mutica*), mixed-mating (*Ae. sharonensis*) and highly selfing (*T. urartu*); same scales for the three species. Each point corresponds to a contig mapped on the *Hordeum vulgare* genome. The blue line is the loess fitting function (degree=2, span=0.1). The dashed black line indicates *H. vulgare* recombination map (in cM/Mb). (B) Mean π_S computed in 50 quantiles of recombination for the three same species. A sigmoid function is fitted (blue line) from which are extracted the slope at the inflection point (dashed line) and the maximum polymorphism level, π_{max}, at the plateau, which is compared to the mean over the whole genome (dotted lined), as exemplified for *Ae. sharonensis.* Note the different y-axis scales. (C) Slope and ratio π_{max}/π_S as a function of PC1 of morphology for the thirteen species, showing that the effect of linked selection strongly increases with selfing. [Figures A and B for all species and chromosomes are given in **Supplementary Fig. S13**].

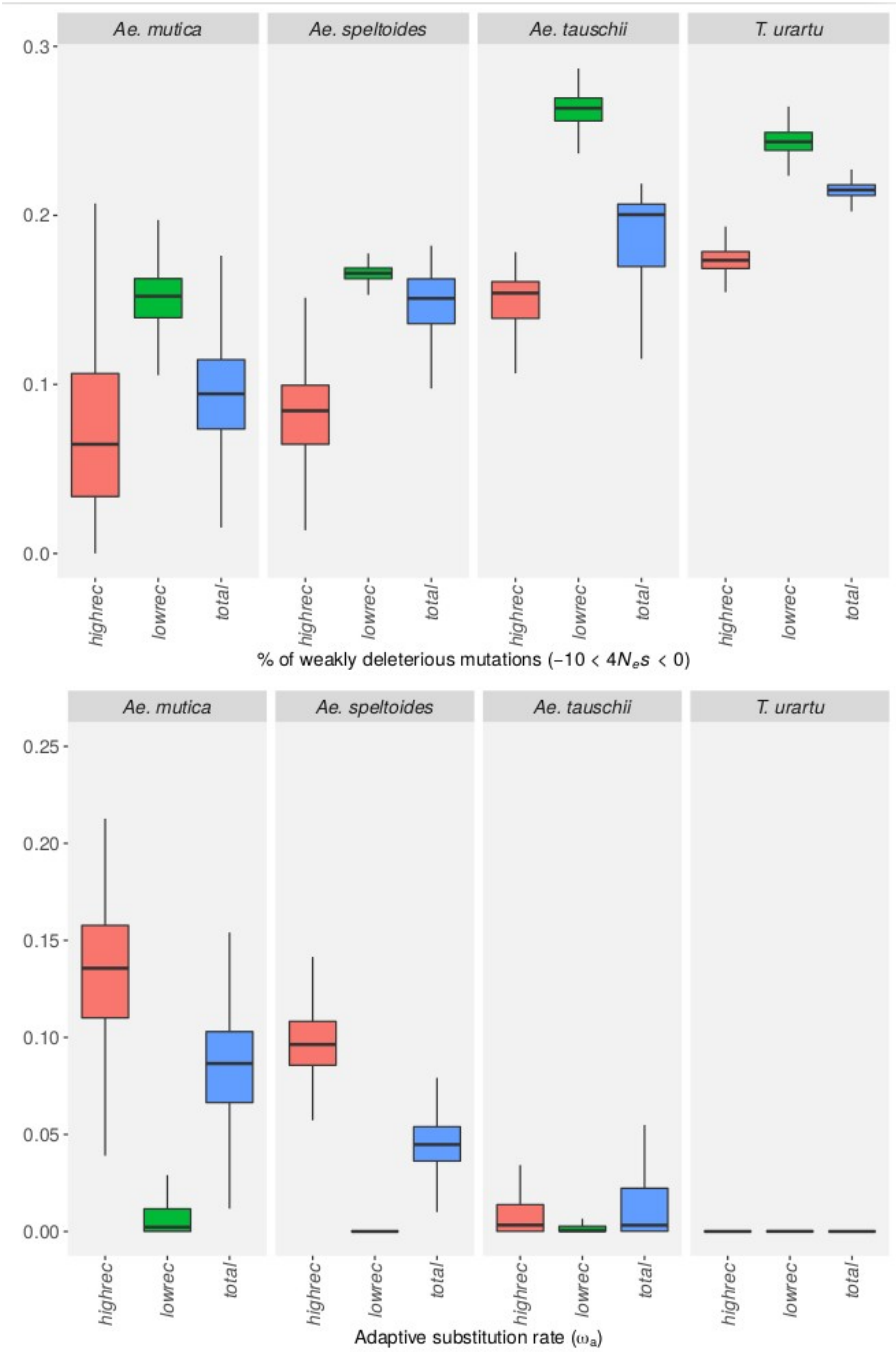

Figure 4. **Mating systems, recombination and the efficacy of selection**. The distribution of fitness effect of mutations was estimated in two self-incompatible species (*Ae. mutica* and *Ae. speltoides*) and in two highly selfing species (*A. tauschii* and *T. urartu*) at the whole genome scale (blue) and for regions of low (green) and high (red) recombination. (A) Proportion of weakly selected deleterious mutations ($-10 < 2N_e s < 0$). (B) Rate of adaptive substitution, i.e. the expected D_N/D_S ratio due to the flux of beneficial mutations.