OPEN

ORIGINAL ARTICLE

The genetic basis and experimental evolution of inbreeding depression in *Caenorhabditis elegans*

IM Chelo[1], S Carvalho[1], M Roque[1], SR Proulx[2] and H Teotónio[1]

Determining the genetic basis of inbreeding depression is important for understanding the role of selection in the evolution of mixed breeding systems. Here, we investigate how androdioecy (a breeding system characterized by partial selfing and outcrossing) and dioecy (characterized by obligatory outcrossing) influence the experimental evolution of inbreeding depression in *Caenorhabditis elegans*. We derived inbred lines from ancestral and evolved populations and found that the dioecious lineages underwent more extinction than androdioecious lineages. For both breeding systems, however, there was selection during inbreeding because the diversity patterns of 337 single-nucleotide polymorphisms (SNPs) among surviving inbred lines deviated from neutral expectations. In parallel, we also followed the evolution of embryo to adult viability, which revealed similar starting levels of inbreeding depression in both breeding systems, but also outbreeding depression. Under androdioecy, diversity at a neutral subset of 134 SNPs correlated well with the viability trajectories, showing that the population genetic structure imposed by partial selfing affected the opportunity for different forms of selection. Our findings suggest that the interplay between the disruptions of coevolved sets of loci by outcrossing, the efficient purging of deleterious recessive alleles with selfing and overdominant selection with outcrossing can help explain mixed breeding systems.
Heredity (2014) **112**, 248–254; doi:10.1038/hdy.2013.100; published online 16 October 2013

INTRODUCTION

Determining the genetics of inbreeding depression is important to understand the role of selection in the maintenance of selfing and outcrossing within populations (Goodwillie *et al.*, 2005; Jarne and Auld, 2006). When inbreeding depression is due to overdominant loci, selfing is disfavored because individuals that self produce more homozygous progeny than if they were to outcross (Ziehe and Roberds, 1989; Charlesworth and Charlesworth, 1990; Uyenoyama and Waller, 1991a). In contrast, selfing is favoured when inbreeding depression is due to deleterious recessive alleles (Lande and Schemske, 1985; Charlesworth *et al.*, 1990; Uyenoyama and Waller, 1991b). Linkage disequilibrium and other non-random associations among deleterious recessives can however result in 'associative' overdominance (Ohta and Kimura, 1970; Ohta, 1971; Palsson and Pamilo, 1999), thus confounding the effects of overdominant loci on the evolution of selfing (Ziehe and Roberds, 1989; David, 1999; Bierne *et al.*, 2000).

Loci underlying differentiation in local environments might also influence the evolution of selfing, especially if they coevolved together within populations and their disruption leads to outbreeding depression (Lynch, 1991; Charlesworth *et al.*, 1997; Epinat and Lenormand, 2009). Selfing in these circumstances might be favoured because it reduces effective segregation and recombination (Nordborg, 2000). However, disruption of coevolved sets of loci can also expose genetic variation at other loci that causes selection for either outcrossing or selfing, depending on whether the newly exposed variation is characterized by overdominance or partial dominance (Nordborg *et al.*, 1996; Schierup and Christiansen, 1996; David, 1999).

There is a poor empirical understanding about whether a balance between different forms of selection can explain mixed breeding systems (Goodwillie *et al.*, 2005; Escobar *et al.*, 2008; Chelo and Teotonio, 2013). To address this problem, we used experimental evolution to study the genetic basis of inbreeding depression in large *Caenorhabditis elegans* populations. These populations were characterized by having either the wild-type male-hermaphrodite androdioecious breeding system (Maupas, 1900; Stewart and Phillips, 2002) or by having a male–female dioecious breeding system (Teotonio *et al.*, 2012). We measured the evolutionary response in the probability of survival with increased inbreeding levels, viability and the diversity of single-nucleotide polymorphisms (SNPs). Our findings suggest that a multistep process involving different forms of selection is responsible for mixed breeding systems.

MATERIALS AND METHODS

Experimental evolution

The construction of the populations and experimental evolution design has been previously detailed (Teotonio *et al.*, 2012). Briefly, the ancestral androdioecious population (termed EEV-A$_0$) resulted from a funnel pairwise cross among 16 wild isolates, while the ancestral dioecious population (EEV-D$_0$) was derived by the recurrent introgression of the *fog-2(q71)* allele into the A$_0$ population for an extra 22 generations. The *fog-2(q71)* allele knocks out self-sperm (Schedl and Kimble, 1988), transforming hermaphrodites into functional females without apparent consequences for male reproductive success (Teotonio *et al.*, 2012).

Ancestral populations defined generation zero (G0) and there was threefold replication for experimental evolution (A$_{1-3}$ and D$_{1-3}$). Populations were cultured alongside for 100 generations at constant 20 °C and 80% relative

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal and [2]Department of Ecology Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, CA, USA
Correspondence: Dr H Teotónio, Instituto Gulbenkian de Ciência, Apartado 14, Oeiras P-2781-901, Portugal.
E-mail: teotonio@igc.gulbenkian.pt

humidity, under discrete 4-day non-overlapping life-cycles at census sizes of $N = 10^4$ (Teotonio et al., 2012). Population samples were periodically stored $-80\,^{\circ}$C.

Inbreeding assays

Population samples were revived from $-80\,^{\circ}$C stocks, each with $>10^3$ individuals, and cultured alongside for two generations under common environmental conditions. In the third generation, L3–L4 larval-staged (immature) individuals were sampled for the inbreeding assays.

Inbred lines of the androdioecious populations were derived by selfing of hermaphrodites for 10 generations, from the A_0 ancestral population, and from each of the three replicate populations at generations 30 and 100. Dioecious inbred lines were derived from G0, G30 and G100 populations by brother–sister mating for 20 generations, to ensure similar final inbreeding coefficients to those of the lines derived from the androdioecious populations (see below).

Lineages were maintained in 12-well culture plates, filled with 3.5 ml of NGM-lite media and 5 µl of O/N cultures of Escherichia coli HT115. L3- or L4-staged individuals were passaged every 4–7 days to new plates. If reproduction or survival of a lineage failed after 7 days, individuals from the previous transfer, kept at $4\,^{\circ}$C, were allowed to reproduce to higher densities and the protocol repeated in the following passage. Extinction of a lineage was scored at the generation where passage was unsuccessful, after three such attempts.

G0 samples were included together with G30 and/or G100 samples in four blocks, defined by the common calendar date of the beginning of the inbreeding protocol. After inbreeding, lines grew to exhaust available food and were further cultured in 9-cm Petri dishes for two generations at high densities and frozen at $-80\,^{\circ}$C. For the ancestral populations over 120 lineages were inbred, for androdioecious evolved replicate populations 72 lineages were inbred and for dioecious evolved replicate populations 48 lineages were inbred (Supplementary Table 1). More derivations were initiated for androdioecious populations because there was little extinction.

Survival analysis

Lineage survival with expected inbreeding coefficients were calculated using a Kaplan–Meier estimator with right-censored data (Therneau and Grambsch, 2000). Expected inbreeding coefficients were defined as: $f_t = 1 - \lambda^t H_0$, where t is generation of selfing or sib mating where extinction was scored, λ a limiting rate quantity set to 0.5 for selfing and to 0.809 for sib mating, and H_0 the average number of heterozygous genotypes before any inbreeding was done, as calculated in each replicate population (Crow and Kimura, 1970). H_0 was previously reported for G0, G30 and G100 population samples, at 334 bi-allelic SNPs measured in chromosomes IV and X (Chelo and Teotonio, 2013).

We tested for the differences in the risk of lineage extinction with f_t either between breeding systems at each generation or between generations within each breeding system. Cox proportional hazards models were employed to calculate different risks of extinction at each breeding system, using block as strata (Therneau and Grambsch, 2000). Ties were handled with the Efron approximation. The formulation using the survival package in the R statistical software (R Development Core Team, 2006; Therneau, 2012): coxph(Surv (ft.extinction, censored.status) ~ strata(block) + mating or generation). Likelihood ratio tests with 1 d.f. were used.

Genotyping of inbred lines

Frozen stocks of G0 and G100 inbred lines were thawed and cultured alongside for two generations at high densities. In the third generation, 20–30 L3- or L4-staged individuals were sampled from each of the inbred lines.

Genomic DNA from pooled individuals was prepared with the ZyGEM prepGEM Insect kit following the manufacturer's protocol (ZyGEM Corporation Ltd, Hamilton, New Zealand). A total of 337 bi-allelic SNPs along chromosomes IV and X were chosen from the genome sequence of the N2, CB4856 and CB4858 wild isolates, as previously described in the study by Chelo and Teotonio (2013). Information about these SNPs is found in Supplementary Table 2. Genotypes were obtained by mass determination, after PCR amplification and allele-specific extension using the iPlex Sequenom MALDITOF platform (Bradic et al., 2011). A total of 26 genotyping runs were

done, each incorporating a maximum of 380 different inbred lines. In each run, 1–4 SNP plexes were used.

Quality control was performed on data including the genotypes from the experimental evolution populations (Chelo and Teotonio, 2013). We first excluded SNPs with $>80\%$ missing data across all samples followed by removal of the inbred lines with $>50\%$ of missing SNP genotypes. After this, SNPs with $>10\%$ of missing data followed by inbred lines with $>10\%$ of missing genotypes were removed.

Physical positions among SNPs were defined according to the C. elegans genome release WS220 (December 2010). Genetic positions among SNPs were obtained by linear interpolation for the two chromosomes, using the function approx in R, each defined with genetic sizes of 50 cM (Rockman and Kruglyak, 2009). Sex determination in C. elegans is chromosomal with hermaphrodites/females XX and males XØ (Hodgkin, 1987). For chromosome IV, SNPs were at densities of 9.4/100 kb (3.3 SNP/cM) and for chromosome X at densities of 9.8/100 kb (3.5 SNP/cM). The number of inbred lines genotyped can be found in Supplementary Table 1 and sample size details per SNP in Supplementary Table 2.

SNP diversity after inbreeding

Genetic diversity among inbred lines was estimated with the previously ascertained SNPs from the study of Chelo and Teotonio (2013). Average effective number of haplotypes was calculated across windows of 10 SNPs with step sizes of 1 SNP along the genetic distance at each chromosome: $h_e = 1/\sum p_i^2$, with p_i being the proportion of haplotype i among inbred lines (Crow and Kimura, 1970). Mean pairwise SNP linkage disequilibrium was also estimated in 10 SNP windows as the composite identity disequilibria among all four genotypes, Δ, assuming that they were the product of the gametic probabilities: $r^2 = \Delta^2/p_a q_a p_b q_b$; with p and q being the proportions of the most and least common allele, respectively, of SNPs a and b (Weir, 1996).

To compute h_e and r^2, SNPs were first phased into haplotypes using fastPHASE 1.2 (Scheet and Stephens, 2006). For each sample of inbred lines 20 random starts of the EM algorithm were employed with 200 haplotypes taken from posterior distributions. The number of clusters for cross-validation was set to 10 and SNPs with posterior probabilities of <0.9 were considered missing data. Note that this protocol accounts for within-population genetic structure and thus the reconstructed inbred lines were diploids that could contain two different haplotypes.

Expected neutral genetic diversity after inbreeding

Monte-Carlo simulations of selfing for 10 generations or full sib mating for 20 generations were performed in order to provide the neutral credible limits on the observed h_e and r^2. A total of 1000 simulations were conducted per replicate population with resulting haplotypes being sampled in the same numbers as the inbred lines. Chromosome IV and chromosome X were analyzed separately. Details on the simulation algorithm can be found in the study by Chelo and Teotonio (2013).

Each run started by randomly sampling phased diploids from the experimental replicates, in equal numbers as those of the starting inbred lineages. Recombination was simulated by exchanging consecutive sets of alleles between the two parental haplotypes (defined as vectors of SNP alleles and ordered as in Supplementary Table 2). We assumed complete crossover interference and map sizes of 50 cM. Crossover occurred randomly between any two consecutive SNPs according to the probability given by the genetic distances between them. For fertilization, and in the case of inbreeding by selfing, two independent gametes were joined to obtain the individual progeny. For brother–sister mating, two genotypes were chosen and kept separately at each generation. For the X chromosome in particular, male genotypes were defined by a single haplotype, to reflect their X-null constitution, by including an extra sampling step after recombination in females.

Viability of ancestral inbred lines

Twenty-five inbred lines from each of A_0 or D_0 populations were randomly revived from $-80\,^{\circ}$C stocks and were cultured alongside for two generations under the same environmental conditions. In the third generation, at day 4 of the life cycle, 100 embryos were collected to 6-cm Petri dish plates, incubated

for 4 days under standard conditions and the number of adult offspring scored after this period. Per inbred line, three assay plates were set up with manipulations and scoring randomized across breeding systems.

Viability was the proportion of adult offspring at each plate out of 100 embryos. Mixed effects ANOVA models were done to estimate the differentiation among breeding systems (Venables and Ripley, 2002). Random inbred lines were modeled within each breeding system and differences estimated by REML with the *lme4* package in R: (lmer(viability ~ mating + (1|line)). For significance, we assumed that the estimated effects followed Student's *t* distributions with 1 d.f.

Viability in outbred populations

Viability assays were carried out in three separate blocks, each including the G0 ancestral populations and one same-numbered replicate population of each breeding system from G10, G20, G41, G70 and G100. For each block, revived population samples were cultured alongside for two generations under common environmental conditions. On the third generation, we set up the assays as above for the G0 inbred lines. Five replicate plates were prepared per population sample.

Differences among breeding systems were modeled at G0 by ANOVA with fixed blocks and fixed breeding systems using the *stats* package in R: (lm(viability ~ block + mating)). Viability trajectories were separately analyzed at each breeding system by fitting mixed effects ANOVAs and estimating differences among generations while considering random block: (lmer(viability ~ generation + (1|block)). For significance, we assumed that the REML estimated effects followed Student's *t* distributions with 1 d.f.

To illustrate the shape of the trajectories we fitted two-segment regressions of the mean viability among replicates onto generation using the *stats* and *segmented* packages in R (Muggeo, 2009): (lm(viability ~ generation); segmented.lm(obj = lm_viability, seg.Z = ~ generation, psi = G40)). To test for the significance of the estimated slopes before and after the break-point Student's *t* distributions were assumed with 1 d.f.

Evolution of SNP diversity in outbred populations

For each replicate experimental population we estimated the mean individual heterozygosity and the deviations in single and multi-locus genotype proportions, from those expected with random mating and infinite sizes (here termed genotype identity disequilibria). All of these metrics were obtained at G0, G10, G30, G70 and G100 from a subset of the 334 SNPs from the study by Chelo and Teotonio (2013), to encompass only those 134 SNPs that are located in the intergenic regions of chromosomes IV and X. They cover 1/3 of the genome and have densities of 1.1 SNP per cM in chromosome IV and 1.6 SNP per cM in chromosome X (Supplementary Table 2). We assume that these SNPs were neutral markers to the loci under putative selection during experimental evolution. For G0, average sample sizes were of ~90 genotypes while for remaining generations average sample sizes were of ~41 genotypes, at each of the SNPs (Supplementary Table 3).

Individual heterozygosity (H_i) is the proportion of heterozygous SNPs across both chromosomes within each individual. H_i is interpreted as the inverse of the expected IBD or the inverse of the expected inbreeding coefficient of a randomly sampled individual; *c.f.* pp.62–68 (Crow and Kimura, 1970). Single-locus genotype disequilibria were measured as the fixation index $F_{IS} = 1 - (H_o/H)$, with H_o being the observed heterozygosity across SNPs and H being the expected heterozygosity under Hardy–Weinberg proportions; *c.f.* pp.104–108 (Crow and Kimura, 1970). For pairwise genotype disequilibria, we calculated the average covariance in pairwise SNP heterozygosity ($g2$), as computed with the RMES software (David et al., 2007; Jarne and David, 2008).

To measure the extent of gametic disequilibria we calculated the 'background' value of r^2. Polynomial functions were first fitted for all the SNPs used in the study by Chelo and Teotonio (2013) against genetic distance, separately at each of the six regions in the chromosomes IV and X that are known to have fairly constant recombination rates, as defined in the study by Rockman and Kruglyak (2009). The genetic distance at which 5% of the initial r^2 decay was reached in each of the six regions was calculated and the average r^2 of the intergenic SNPs pairwise combinations above this distance taken as the background LD (bkgLD). bkgLD is inversely correlated to effective recombination rates, as r^2 is a function of gamete combinations (see above) and because selection should not distort genotype disequilibria among loci separated by large genetic distances.

Correlations of viability with SNP diversity

For selection to maintain excess diversity, as previously measured in the experimental outbred populations (Chelo and Teotonio, 2013), the mean and/or the variance among individual inbreeding coefficients and genotype identity disequilibria must be associated with fitness (David, 1999; Bierne et al., 2000; Navarro and Barton, 2002; Szulkin et al., 2010). To illustrate these associations, the Pearson product moment correlations of viability with H_i, F_{IS}, $g2$ and bkgLD were calculated using the means of the three replicate populations within each breeding system at G10, G30, G70 and G100. Viability data for G20 and G41 were averaged per replicate in order to pair it with the SNP data at G30. For significance testing, Pearson coefficients were transformed to Fisher's *z*-coefficients and across all generations one-tailed *t*-tests with 3 d.f. were performed. Spearman coefficients gave similar results (analysis not shown).

RESULTS

Survival upon inbreeding

The proportion of surviving lineages with inbreeding by selfing in androdioecious populations or with sib mating in dioecious populations is shown in Figure 1 (see also Supplementary Figure 1). In the ancestral generation, the survival rate of dioecious lineages was ~0.6 after inbreeding (inbreeding coefficients of $f_t \approx 1$), corresponding to a 2.5-fold increase in the probability of extinction over that of the androdioecious populations (LK ratio test = 72.1, $\log_{10} P = -7.23$, $n = 255$). These differences continued to be apparent during experimental evolution. At G30, dioecious populations had a twofold higher chance of going extinct when inbred than androdioecious populations
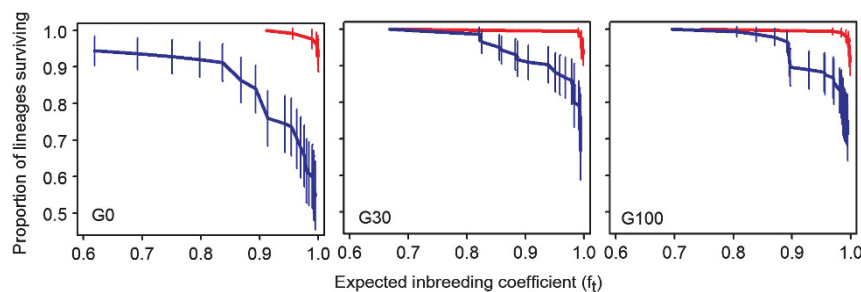

Figure 1 Survival rates with inbreeding. Proportion of lineages surviving multiple generations of inbreeding by selfing (red) or sib mating (blue) in the androdioecious or dioecious populations, respectively, at different generations of experimental evolution (G0, G30 and G100). Error bars show 2 × s.d. See also Supplementary Figure 1.

(LK ratio test $= 56.8$, $\log_{10} P = -13.3$, $n = 360$), a difference that by G100 was reduced to a 1.5-fold higher extinction risk (LK ratio test $= 45.1$, $\log_{10} P = -10.7$, $n = 360$).

Analysis of lineage survival at each breeding system across the three periods indicated no evolution under androdioecy (LK ratio test $= 2.1$, $P = 0.148$, $n = 562$) and a marginal increase during evolution under dioecy (LK ratio test $= 3.5$, $P = 0.063$, $n = 413$).

SNP diversity after inbreeding

At G0, SNP diversity among the inbred lines deviated from that expected with neutral processes during inbreeding (Figure 2). In both androdioecious and dioecious ancestors, inbred lines had higher haplotype diversity (h_e) at chromosome IV than expected (Figure 2a). For chromosome X, however, there was no excess diversity (Figure 2b).

By G100, selfing androdioecious hermaphrodites resulted in increased h_e in both chromosomes relative to neutrality and when compared with the diversity of the outbred populations from which the inbred lines were derived (Figures 2a and b). Inbreeding the dioecious populations achieved higher h_e than neutral expectations but the inbred lines showed reduced diversity relative to that of the outbred populations. Further, in contrast to androdioecy, under dioecy higher h_e was only apparent in the autosome.

Linkage disequilibrium among the inbred lines (r^2) generally followed neutral expectations despite breeding system, chromosome and generation of experimental evolution (Figure 2c and d). However, and particularly for G100, all measured r^2 tended to cluster by the lower credible limits of neutrality. In fact, two out of three replicate androdioecious populations had significantly lower r^2 than expected with neutrality.
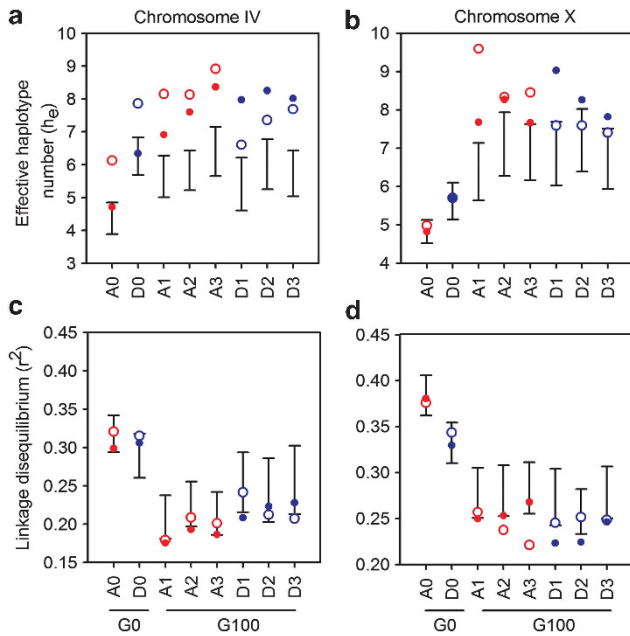
Evolution of viability

Before inbreeding, the ancestral dioecious population was 8% less viable than the ancestral androdioecious population (Figure 3a; $F_{1,26} = 9.89$, $P = 0.004$; block n.s.). Both selfing and sib mating led to 13% reductions in the mean values observed among inbred lines, with androdioecious lines being 8% more viable than dioecious lines ($t = -4.7$, $P < 0.001$).

There was evidence for the experimental evolution of viability only under dioecy. For dioecy, regression analysis showed that the break-point is at G60 (60.44 ± 8.61 s.d.; adj. $R^2 = 81\%$), and both before and after slopes are significant (pre-G60 $t = 3.49$, $P < 0.001$; post-G60 $t = -4.08$, $P < 0.001$). For androdioecy, the break-point is at G20 (19.14 ± 17.53 SD) although the model has a very poor fit (adj. $R^2 = 10\%$) and neither slope is significant (pre-G20 $t = -0.728$, $P = 0.466$; post-G20 $t = 0.910$, $P = 0.363$).

Correlated evolution of viability with SNP diversity

The correlations of SNP diversity with viability were most obvious under androdioecy (Figure 4). In this breeding system, tests across all generations showed a significant correlation of viability with $bkgLD$ ($P = 0.02$) and marginal significance with $g2$ ($P = 0.06$). When not including G100, all androdioecious correlations are different from zero ($P = 0.01$). In contrast to androdioecy, under dioecy, tests across all generations only revealed significance of the correlation of $g2$ with viability ($P = 0.04$).

DISCUSSION

Inbreeding and outbreeding depression

Inbreeding depression is thought to mostly occur because deleterious recessive alleles are expressed in homozygotes (Charlesworth and
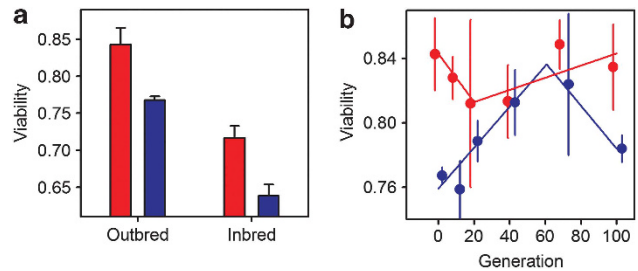

Figure 2 Genetic diversity after inbreeding. Effective haplotype number (**a**, **b**) and linkage disequibrium (**c**, **d**) for chromosomes IV (**a**, **c**) and chromosomes X (**b**, **d**). Circles show the diversity measured after inbreeding among the androdioecious lines (red) or dioecious lines (blue). Points indicate the diversity of the experimental populations before inbreeding. Error bars show the 95% credible probability obtained with 1000 neutral simulations of inbreeding.

Figure 3 Evolution of viability. (**a**) Outbred or inbred viability of the ancestral androdioecious population (red) and the ancestral dioecious population (blue). Error bars show 1 s.e.m. among assay blocks for the experimental populations or one s.e.m. among the inbred lines. (**b**) Viability of androdioecious (red) and of dioecious (blue) populations during experimental evolution. Error bars indicate 1 s.e.m. among the three replicates. Lines show the estimated two-segment regressions.
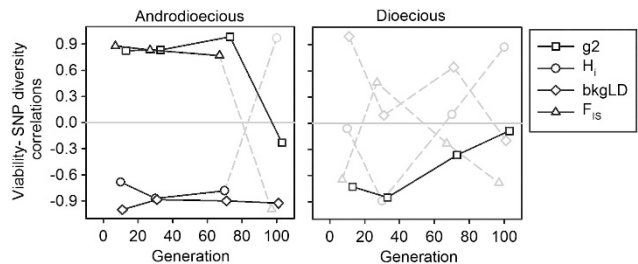

Figure 4 Population structure and selection. Correlation coefficients of viability with several genotype disequilibria metrics during experimental evolution. In solid lines, comparisons that are significantly different from zero, gray dashed lines otherwise.

Willis, 2009). As selection against recessive deleterious alleles is weaker under outcrossing than under selfing, we expected that inbreeding depression would be maintained at higher levels under dioecy than under androdioecy (Lande and Schemske, 1985; Charlesworth et al., 1990; Uyenoyama and Waller, 1991b). As expected, we found that dioecious populations subjected to inbreeding had low lineage survival at all generations of experimental evolution, whereas androdioecious populations subjected to inbreeding had high lineage survival rates.

The viability results in the ancestral populations are also consistent with inbreeding depression being due to deleterious recessives because the inbred lines were less viable than the outbred populations from which they were derived. Further, the viability data showed initial outbreeding depression, with dioecious populations having lower viability than androdioecious populations, regardless of the level of inbreeding. Interestingly, therefore, outbreeding and inbreeding depression were not independent phenomena; c.f., (Lynch, 1991; Schierup and Christiansen, 1996; Escobar et al., 2008; Epinat and Lenormand, 2009).

Outbreeding depression implies underdominance (Lynch, 1991), following the disruption by segregation and recombination of sets of loci that coevolved in different populations through adaptation to local environmental conditions or by genetic drift (Coyne and Orr, 1998; Kirkpatrick and Barton, 2006; Epinat and Lenormand, 2009). However, underdominance would not have led to the excess diversity that we measured among the inbred lines. Instead, there must have been fitness overdominance during inbreeding. Notably, a similar conclusion was reached when excess diversity relative to neutral expectations was detected in the experimental outbred populations (Chelo and Teotonio, 2013).

It is known that fitness overdominance can result from the non-random association of deleterious recessive alleles that are not necessarily in close physical linkage (Ohta and Kimura, 1970; Ohta, 1971; Palsson and Pamilo, 1999). Could this associative overdominance help explain the interdependence of outbreeding and inbreeding depression during experimental evolution? Specifically, as long as the coevolved sets of loci remained intact there would be outbreeding depression but once disrupted would there be inbreeding depression? An answer to this question needs a better understanding of how the population genetic structure imposed by the two breeding systems influences the opportunity for different forms of selection (David, 1999; Szulkin et al., 2010).

Population structure and selection
As expected with little population genetic structure (Charlesworth et al., 1990; Szulkin et al., 2010), under dioecy there was no trend in the correlations of SNP diversity with viability over the course of the experimental evolution, implying that selection was not very efficient at removing deleterious recessive alleles and/or at sustaining fitness overdominance.

In the androdioecious populations, however, several correlations were significant. There was a negative correlation between viability with background linkage disequilibrium (bkgLD), which suggests that selection favoured new recombinants. This is because bkgLD should quantify the extent of gametic linkage disequilibrium and thus effective recombination (Christiansen, 1989). Additionally, the correlations between the co-variation in diversity within SNPs (F_{IS}) and the co-variation in diversity among SNPs (g2) with viability were positive. Positive signs in these correlations indicate that identity disequilibria at multiple loci across the genome facilitated selection among genetically heterogeneous individuals (David, 1999; Szulkin

et al., 2010). This population genetic structure may have in turn reinforced selection against deleterious recessive alleles, which is consistent with the observation of a negative correlation of viability with individual heterozygosity (H_i).

If the population genetic structure imposed by partial selfing in androdioecy enabled the purging of deleterious recessive alleles, then the fitness overdominance responsible for excess diversity in the outbred populations might have been due to truly overdominant loci (Christiansen, 1989; Ziehe and Roberds, 1989; Charlesworth and Charlesworth, 1990). In fact, the positive correlations between viability and F_{IS} and g2 found under androdioecy suggest selection on overdominant loci. This is because these metrics also quantify the number of heterozygote classes (Weir et al., 1980; David, 1999; David et al., 2007), and it has been theoretically shown that selection on a few overdominant loci creates positive correlations between the number of heterozygote classes and fitness variance (David, 1999). Furthermore, if the average population fitness was a diminishing returns function of heterozygosity the negative correlations of viability with bkgLD under androdioecy could be generated by overdominant loci (Navarro and Barton, 2002).

Reductions in viability with inbreeding were accompanied by higher diversity among inbred lines than among the individuals of the outbred populations from which they derived, a pattern particularly evident in the autosomes at the beginning of experimental evolution. These results can be explained by the generation of new deleterious alleles during inbreeding, of which only some were selected against, c.f. (Barriere et al., 2009). Diversity results among inbred lines at generation 100 further support a role for truly overdominant loci. Dioecious sib mating led to less diversity among inbred lines than selfing under androdioecy. Despite the possibility for similar kinds of selection in the two breeding systems, upon inbreeding, reduced effective recombination with selfing could have resulted in higher (haplotype) diversity because of selection on overdominant loci (Navarro and Barton, 2002).

Selection and the maintenance of androdioecy
When sexual selection is considered the diversity results can be more fully explained (Anthes et al., 2010; Baer et al., 2010; Mallet and Chippindale, 2011). In particular, sexual selection should have been stronger under dioecy because higher numbers of males in this breeding system would have led to lower numbers of X chromosomes—in C. elegans males are XØ and hermaphrodites/females are XX (Hodgkin, 1987). Following inbreeding there was excess diversity in the X-chromosome under androdioecy but not under dioecy. As expected, therefore, the sex ratio of a population might have been associated with the removal of deleterious recessive alleles from the X-chromosome. This result is remarkable because we were previously unable to confirm that sex ratio differences among breeding systems influenced the evolution of male competitive performance (Teotonio et al., 2012), presumably a fitness component under strong sexual selection, compare with (LaMunyon and Ward, 2002; Murray et al., 2011).

Taken together, our findings point to a multistep process that maintains partial selfing under androdioecy (also refer to Charlesworth et al., 1997; Pannell, 2002; Goodwillie et al., 2005). Transitions from outcrossing to selfing can first enable the appearance of different sets of coevolved loci, given sufficient time for differentiation among populations. These transitions to selfing could occur, for example, because of reproductive assurance during the colonization of empty habitats (Cheptou, 2004). However, some degree of outcrossing is inevitable because of recurrent mutation in sex

determination pathways and dispersal of males among neighboring populations. The disruption of coevolved sets of loci would in turn expose to selection partially dominant loci that originated during population differentiation, which would favor selfing. Because of purging deleterious recessives, however, transient associative fitness overdominance and/or selection on truly overdominant loci would allow outcrossing to persist until local population extinction and novel transitions to selfing.

Several lines of evidence suggest the occurrence of this multistep process in the maintenance of androdioecy in *C. elegans*. Hybridization of wild isolates results in outbreeding depression that might be due to several loci in complete linkage disequilibrium (Dolgin *et al.*, 2007; Seidel *et al.*, 2008). There is also abundant genetic variation for male function and selection for outcrossing in novel environments (Teotonio *et al.*, 2006; Manoel *et al.*, 2007; Murray *et al.*, 2011; Teotonio *et al.*, 2012), even if males are rarely found in natural populations (Felix and Duveau, 2012). Finally, heterozygosity within natural populations might be higher than that expected with a long history of exclusive selfing (but see Barriere and Felix, 2005; Sivasundar and Hey, 2005; Cutter *et al.*, 2008; Andersen *et al.*, 2012). All these observations indicate that in *C. elegans* a balance between different forms of selection maintains both selfing and outcrossing. Perhaps a similar balance also explains mixed breeding systems in other species.

DATA ARCHIVING

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS et al. (2012). Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet* **44**: 285–290.

Anthes N, David P, Auld JR, Hoffer JN, Jarne P, Koene JM et al. (2010). Bateman gradients in hermaphrodites: an extended approach to quantify sexual selection. *Am Nat* **176**: 249–263.

Baer CF, Joyner-Matos J, Ostrow D, Grigaltchik V, Salomon MP, Upadhyay A (2010). Rapid decline in fitness of mutation accumulation lines of gonochoristic (outcrossing) Caenorhabditis nematodes. *Evolution* **64**: 3242–3253.

Barriere A, Felix MA (2005). High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr Biol* **15**: 1176–1184.

Barriere A, Yang SP, Pekarek E, Thomas CG, Haag ES, Ruvinsky I (2009). Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. *Genome Res* **19**: 470–480.

Bierne N, Tsitrone A, David P (2000). An inbreeding model of associative overdominance during a population bottleneck. *Genetics* **155**: 1981–1990.

Bradic M, Costa J, Chelo IM (2011). Genotyping with Sequenom. In: Orgogozo V, Rockman M (eds) *Molecular Methods for Evolutionary Genetics*, Vol. 772. Humana Press: New York, NY, USA.

Charlesworth B, Nordborg M, Charlesworth D (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* **70**: 155–174.

Charlesworth D, Charlesworth B (1990). Inbreeding depression with heterozygote advantage and its effect on selection for modifiers changing the outcrossing rate. *Evolution* **44**: 870–888.

Charlesworth D, Morgan MT, Charlesworth B (1990). Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**: 1469–1489.

Charlesworth D, Willis JH (2009). The genetics of inbreeding depression. *Nat Rev Genet* **10**: 783–796.

Chelo IM, Teotonio H (2013). The opportunity for balancing selection in experimental populations of *Caenorhabditis elegans*. *Evolution* **67**: 142–156.

Cheptou PO (2004). Allee effect and self-fertilization in hermaphrodites: reproductive assurance in demographically stable populations. *Evolution* **58**: 2613–2621.

Christiansen FB (1989). Linkage equilibrium in multi-locus genotypic frequencies with mixed selfing and random mating. *Theor Pop Biol* **35**: 307–336.

Coyne JA, Orr HA (1998). The evolutionary genetics of speciation. *Philos Trans R Soc Lond Ser B Biol Sci* **353**: 287–305.

Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper & Row, Publishers: New York, NY, USA.

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

David P (1999). A quantitative model of the relationship between phenotypic variance and heterozygosity at marker loci under partial selfing. *Genetics* **153**: 1463–1474.

David P, Pujol B, Viard F, Castella V, Goudet J (2007). Reliable selfing rate estimates from imperfect population genetic data. *Mol Ecol* **16**: 2474–2487.

Dolgin ES, Charlesworth B, Baird SE, Cutter AD (2007). Inbreeding and outbreeding depression in Caenorhabditis nematodes. *Evolution* **61**: 1339–1352.

Epinat G, Lenormand T (2009). The evolution of assortative mating and selfing with in- and outbreeding depression. *Evolution* **63**: 2047–2060.

Escobar JS, Nicot A, David P (2008). The different sources of variation in inbreeding depression, heterosis and outbreeding depression in a metapopulation of Physa acuta. *Genetics* **180**: 1593–1608.

Felix MA, Duveau F (2012). Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. *BMC Biol* **10**: 59.

Goodwillie C, Kalisz S, Eckert CG (2005). The evolutionary enigma of mixed mating systems in plants: occurrence, theoreticalexplanations, and empirical evidence. *Ann Rev Ecol Evol Syst* **36**: 47–79.

Hodgkin J (1987). Sex determination and dosage compensation in *Caenorhabditis elegans*. *Ann Rev Genet* **21**: 133–154.

Jarne P, Auld JR (2006). Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution* **60**: 1816–1824.

Jarne P, David P (2008). Quantifying inbreeding in natural populations of hermaphroditic organisms. *Heredity* **100**: 431–439.

Kirkpatrick M, Barton N (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

LaMunyon CW, Ward S (2002). Evolution of larger sperm in response to experimentally increased sperm competition in *Caenorhabditis elegans*. *Proc Biol Sci R Soc* **269**: 1125–1128.

Lande R, Schemske DW (1985). The evolution of inbreeding depression and selfing in plants. I. Genetic models. *Evolution* **39**: 24–40.

Lynch M (1991). The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* **45**: 622–629.

Mallet MA, Chippindale AK (2011). Inbreeding reveals stronger net selection on *Drosophila melanogaster* males: implications for mutation load and the fitness of sexual females. *Heredity* **106**: 994–1002.

Manoel D, Carvalho S, Phillips PC, Teotonio H (2007). Selection against males in *Caenorhabditis elegans* under two mutational treatments. *Proc Biol Sci R Soc* **274**: 417–424.

Maupas E (1900). Modes et formes de reproduction des nematodes. *Arch Exp Gen Ser* **3**: 463–624.

Muggeo VMR (2009). segmented: Segmented relationships in regression models. *R Package Version* 02-5.

Murray RL, Kozlowska JL, Cutter AD (2011). Heritable determinants of male fertilization success in the nematode *Caenorhabditis elegans*. *BMC Evol Biol* **11**: 99.

Navarro A, Barton NH (2002). The effects of multilocus balancing selection on neutral variability. *Genetics* **161**: 849–863.

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

Nordborg M, Charlesworth B, Charlesworth D (1996). Increased levels of polymorphim surrounding selectively maintained sites in highly selfing species. *Proc Biol Sci R Soc* **263**: 1033–1039.

Ohta T (1971). Associative overdominance caused by linked detrimental mutations. *Genet Res* **18**: 277–286.

Ohta T, Kimura M (1970). Development of associative overdominance through linkage disequilibrium in finite populations. *Genet Res* **16**: 165–177.

Palsson S, Pamilo P (1999). The effects of deleterious mutations on linked, neutral variation in small populations. *Genetics* **153**: 475–483.

Pannell JR (2002). The evolution and maintenance of androdioecy. *Ann Rev Ecol Syst* **33**: 397–425.

R Development Core Team (2006). R: a language and environment for statistical computing. http://www.R-project.orgVienna

Rockman MV, Kruglyak L (2009). Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet* **5**: e1000419.

Schedl T, Kimble J (1988). fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. *Genetics* **119**: 43–61.

Scheet P, Stephens M (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**: 629–644.

Schierup MH, Christiansen FB (1996). Inbreeding and outbreeding depression in plants. *Heredity* **77**: 461–468.

Seidel HS, Rockman MV, Kruglyak L (2008). Widespread genetic incompatibility in C. elegans maintained by balancing selection. *Science* **319**: 589–594.

Sivasundar A, Hey J (2005). Sampling from natural populations with RNAI reveals high outcrossing and population structure in *Caenorhabditis elegans*. *Curr Biol* **15**: 1598–1602.

Stewart AD, Phillips PC (2002). Selection and maintenance of androdioecy in *Caenorhabditis elegans*. *Genetics* **160**: 975–982.

Szulkin M, Bierne N, David P (2010). Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64**: 1202–1217.

Teotonio H, Carvalho S, Manoel D, Roque M, Chelo IM (2012). Evolution of outcrossing in experimental populations of *Caenorhabditis elegans*. *PLoS One* **7**: e35811.

Teotonio H, Manoel D, Phillips PC (2006). Genetic variation for outcrossing among *Caenorhabditis elegans* isolates. *Evolution* **60**: 1300–1305.

Therneau T (2012). A package for survival analysis in R. *R Package Version* 236-4.

Therneau TM, Grambsch PM (2000). *Modeling Survival Data*. Springer: New York, USA.

Uyenoyama MK, Waller DM (1991a). Coevolution of self-fertilization and inbreeding depression. II. Symmetric overdominance in viability. *Theor Popul Biol* **40**: 47–77.

Uyenoyama MK, Waller DM (1991b). Coevolution of self-fertilization and inbreeding depression. III. Homozygous lethal mutations at multiple loci. *Theor Popul Biol* **40**: 173–210.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. Springer: New York, NY, USA.

Weir BS (1996). *Genetic Data Analysis II*. Sinauer Associates, Inc: Sunderland, MA, USA.

Weir BS, Avery PJ, Hill WG (1980). Effect of mating structure on variation in inbreeding. *Theor Pop Biol* **18**: 396–429.

Ziehe M, Roberds JH (1989). Inbreeding depression due to overdominance in partially self-fertilizing plant populations. *Genetics* **121**: 861–868.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)