Temporal Dynamics and Linkage Disequilibrium in Natural C. elegans Populations.

Antoine Barrière, Marie-Anne Félix

This manuscript has been accepted for publication in Genetics (2007) doi : 10.1534/genetics.106.067223

Temporal Dynamics and Linkage Disequilibrium in Natural *C. elegans* Populations

Antoine Barrière[1] and Marie-Anne Félix
Institut Jacques Monod, CNRS – Universities of Paris 6 and 7, 75251 Paris cedex 05, France.
e-mail: felix@ijm.jussieu.fr

[1]present address: Department of Ecology and Evolution, the University of Chicago, Chicago, IL 60637 USA

A. B. dedicates this work to the memory of Daniel Lachaise, who had a decisive influence on his career.

Running title: Temporal Dynamics of *C. elegans*

Keywords: population genetics, Caenorhabditis elegans, temporal survey, outbreeding depression

Corresponding author: Marie-Anne Félix

Institut Jacques Monod, CNRS – Universities of Paris 6 and 7, Tour 43, 2 place Jussieu, 75251 Paris cedex 05, France. Phone: +33 1 44 27 40 88, Fax: +33 1 44 27 52 65, e-mail: felix@ijm.jussieu.fr

Summary:

Caenorhabditis elegans is a major laboratory model system yet a newcomer to the field of population genetics, and relatively little is known of its biology in the wild. Recent studies of natural populations at a single timepoint revealed strong spatial population structure and suggested that these populations may be very dynamic. We have therefore studied several natural *C. elegans* populations over time and genotyped them at polymorphic microsatellite loci. While some populations appear to be genetically stable over the course of observation, others seem to go extinct, with full replacement of multilocus genotypes upon regrowth. The frequency of heterozygotes indicates that outcrossing occurs at a mean frequency of 1.7% and is variable between populations. However, in genetically stable populations, linkage disequilibrium between different chromosomes can be maintained over several years, at a level much higher than expected from the heterozygote frequency. *C. elegans* seems to follow metapopulation dynamics, and the maintenance of linkage disequilibrium despite a low yet significant level of outcrossing suggests that selection may act against the progeny of outcrossings.

INTRODUCTION

Most population genetic studies infer evolutionary mechanisms of a population from a single timepoint. There are as yet few molecular studies of the same population over time (VIARD *et al.* 1997; GUILLEMAUD *et al.* 2003; MEUNIER *et al.* 2004; CHARBONNEL and PEMBERTON 2005; TROUVÉ *et al.* 2005). However, to get a direct picture of dynamic phenomena such as migration, recombination, selection or population extinction, and to detect variation over time, temporal surveys of populations are required. Here we present a temporal study of several natural populations of the nematode *Caenorhabditis elegans* over three years.

Beyond the fact that *C. elegans* is a major laboratory model organism with a fast generation time (3.5 days in standard laboratory conditions), an interesting feature for evolutionary biology is its peculiar mode of reproduction: *C. elegans* has two sexes, selfing XX hermaphrodites and facultative XO males that are able to mate with hermaphrodites. Males arise either spontaneously by rare non-disjunction of X chromosomes at meiosis (at a rate around 0.1%; HODGKIN and DONIACH 1997; TEOTÓNIO *et al.* 2006) or as progeny of hermaphrodites when mated with males (50% of the cross progeny is male). This facultative outcrossing makes *C. elegans* an excellent system to study the impact of outcrossing in a diploid organism.

Studies on natural populations of *C. elegans* have only recently begun. This species displays low overall levels of polymorphism (similar to humans, but 20-fold lower than *Drosophila melanogaster*) and displays only weak geographic structure at a worldwide scale (KOCH *et al.* 2000; DENVER *et al.* 2003; BARRIÈRE and FÉLIX 2005; CUTTER 2006, but see HABER *et al.* 2005). Selfing is clearly the predominant mode of reproduction in the wild, but outcrossing rate estimates range from 0.01%, to 1-20%. The lower number (0.01%) was estimated from studies of linkage disequilibrium, either within local populations (BARRIÈRE and FÉLIX 2005) or among a worldwide set of isolates (CUTTER 2006). The higher range was

estimated from measures of heterozygote frequencies in populations from France (1.3%; BARRIÈRE and FÉLIX 2005) and Los Angeles (20%; SIVASUNDAR and HEY 2005).

The habitat of *C. elegans* on ephemeral resources and its population genetic structure led to the suggestion that it may follow metapopulation dynamics (BARRIÈRE and FÉLIX 2005; SIVASUNDAR and HEY 2005), with populations frequently going extinct and habitats being recolonized through migration (HANSKI 1999). However, this was suggested by indirect evidence, and a temporal study of local populations has so far been lacking. We have thus followed *C. elegans* populations in several locations over the span of one to three years, with the goal of identifying the population dynamics shaping its evolution. We found that some *C. elegans* populations were ephemeral, and observed metapopulation dynamics, with extinction followed by recolonization by new genotypes. Surprisingly, in the largest and most stable population, genetic linkage between the same alleles persisted over three years despite a detectable rate of outcrossing, suggesting selection acting against the progeny of a recombination event. The observed metapopulation structure and the maintenance of linkage disequilibrium may explain the discrepancy between outcrossing rates measured at short and long timescales through heterozygote frequency and linkage disequilibrium, respectively. We finally discuss how the observed dynamics of natural *C. elegans* populations may influence the genetic and phenotypic evolution of this species.

MATERIALS AND METHODS

Sampling

The sampling procedure and most locations were described in BARRIÈRE and FÉLIX (2005). New locations include Obernai, Bas-Rhin, France (position: 48.46°N, 7.48°E), leaf litter next to a vegetable garden in the middle of vineyards, and the Botanical garden of the

University of Lisbon, Portugal (38.42°N, 9.12°W), leaf and fruit litter below trees and in a compost heap. Samples were collected from September 2002 to January 2006 (details on sampling can be found in Table S1).

Sampled material was spread on standard NGM plates seeded with *E. coli* OP50. Worms were picked within 1 hour to 2 days after plating; the developmental stage was recorded, as described in BARRIÈRE and FÉLIX (2006). On several occasions, six samples were taken a few centimeters apart in a single compost pile.

In most cases, individuals sampled from the soil were left to develop on the plates and self-progeny of hermaphrodites were harvested. One portion was frozen in glycerol and kept at -80°C while the rest was lysed in Worm Lysis Buffer (a digestion buffer commonly used to prepare DNA from *C. elegans*: 50 mM KCl, 10 mM Tris pH 8.2, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, 0.01% gelatin, 0.06% proteinase K), the lysate conserved at -20°C and used for subsequent amplification. This procedure conserves the genotype of the collected individual, even when heterozygous (labeled 'heterozygous' in Table 2). In other cases (labeled 'inbred' in Table 2), isogenic strains were established by selfing for several generations (BARRIÈRE and FÉLIX 2005), and possible heterozygosity was lost.

Microsatellite genotyping

Six microsatellite regions defined in HABER *et al.* (2005) were amplified by PCR. The forward primer was labeled with a fluorophore, either Hex or 6-Fam, and amplified fragments were run on a ABI 3100-avant system. Primers were:

II-L: f: AACAAAAATGTGGCAGGGAG, r: GGGTTACGGTAGTGGTACTGTAGG.

III-R: f: GATGAATGGATATGACCGGC, r: TATCAGGCGTATCACCTCCC.

IV-L: f: AAGATTTCTGCTAACGTGCTGA, r: AGTAACTTTGGTGCAGGTTCG.

V-L: f: CGTTGGGACAGGATCTAGTTG, r: CGTTGGGACAGGATCTAGTTG.

X-R: f: GCACACGCTTGAATGTCATAA, r: AAGAGCAGTAGCCGTTGTTGA.

For the *II-R* locus, we used a slightly different protocol, with the forward primer tailed with an M13 sequence, and amplification being conducted with a labeled M13 primer. Primers were: f: CACGACGTTGTAAAACGACTTCTCATTGGAAAGTTGGGC, r: CAATACCGAGAAACGGATGAA.

When a putative heterozygote was found, the glycerol stock was thawed, revived and individual worms were genotyped to check whether both alleles segregated. Microsatellite repeat numbers were deduced from PCR fragment length and comparison with the repeat number in N2.

For microsatellite locus *II-L* in samples from Le Perreux-sur-Marne, we had reproducibility problems, and therefore excluded this locus from further analysis in this population. In several isolates from the Lis12-0705 sample, genotyping at locus *IV-L* revealed two different fragment sizes, corresponding to 35 and 44 repeats. This polymorphism did not segregate in the self progeny as would be expected from a heterozygote at a single locus, and the progeny of a cross with N2 males displayed three allele sizes (that of N2 plus the two others from Lis12-0705), suggesting a duplication of this locus in Lis12-0705. Since one individual (LisbonP12D3, Dataset S1) displayed only the 35-repeat allele, we considered this 35-repeat allele to be at the locus *IV-L* genotyped in other strains. Similarly, one individual from the Lis14-0705 sample displayed two fragment sizes, corresponding to 36 and 46 repeats; we considered the 36-repeat allele to be at the locus orthologous to *IV-L.*

Data analysis

The pairwise difference, the scaled mutation parameter Θ_{hom} and gene diversity *H* were calculated with Arlequin V. 3.01 (EXCOFFIER *et al.* 2005), over all loci. Θ_{hom} derives from gene diversity by the relationship $H=\frac{\Theta_{hom}}{\Theta_{hom}+1}$. Population structure, as measured by θ

(WEIR and COCKERHAM 1984), an estimator of F_{ST}, and 95% confidence intervals were calculated by bootstrap with FSTAT (GOUDET 2001).

The best statistics to test for population differentiation with unbalanced samplings is not F_{ST} or its components, but the likelihood ratio G statistic (GOUDET *et al.* 1996). Differentiation between samples was tested in R (R DEVELOPMENT CORE TEAM 2003) with the package hierfstat (GOUDET 2005).

To calculate the inbreeding coefficient f, (WEIR and COCKERHAM 1984), an estimator of F_{IS}, we used GDA (LEWIS and ZAYKIN 2001). Its confidence interval was obtained by bootstrapping over loci. Selfing rate was calculated as $s=2f/(1+f)$.

Genotypes were obtained either on the pooled progeny of an individual sampled from the wild ('non-inbred'), or after inbreeding for a few generations in the laboratory by picking a single hermaphrodite individual ('inbred' strains, designated with a JU strain number). This feature is indicated for each sample in the second column of Table 2. For calculations involving comparisons between laboratory inbred and non-inbred genotypes, we considered inbred strains as non-inbred that are homozygotes at all loci: given the very high level of inbreeding witnessed in non-laboratory-inbred populations, this can be considered a reasonably good approximation.

In the case of non-inbred genotypes, we deduced haplotypes from genotypes: in most cases, they were homozygous; for the five individuals that were heterozygous at two or more loci, we inferred the phase from other haplotypes found in the same sample.

The standardized index of association I_A^S (multilocus measure of linkage disequilibrium) and its significance (p-value) were calculated for the different samples with Lian v. 3.5 (HAUBOLD and HUDSON 2000), using the parametric test.

For the confidence intervals on linkage disequilibrium in Franconville, values of D' were calculated from data, and the expected genotype frequency was calculated, based on

allele frequencies, for gradually more distant values of D'. The concordance of these expected tables with the observed data was then calculated by a polynomial probability estimate. Pairwise comparisons between two timepoints were carried with the same method, calculating the concordance of D' values between the two samples. The R function used is available from the authors at request.

For estimations of generation times compatible with the observed linkage disequilibrium over time, the latter was considered to decay according to $D'_{N+1}=(1-r)D'_N$, N being the generation number, r being the recombination rate. At generation N, linkage disequilibrium would be $D'_N=(1-r)^N D'_0$. From linkage disequilibrium, it is thus possible to calculate N as $N=\ln(D'_N/D'_0)/\ln(1-r)$. However, in *C. elegans*, r is diminished by inbreeding and lower than the normal recombination rate. Therefore, the observed recombination rate is $r'=r(1-F_{IS})$ (NORDBORG 1997). Thus, $N=\ln(D'_N/D'_0)/\ln(1-r(1-F_{IS}))$.

Embryonic lethality

To check for possible incompatibility between genotypes, embryonic lethality and brood size were monitored in the F2 progeny of interstrain crosses (JU360 males with JU361 hermaphrodites, and the reverse cross). F1 hermaphrodites were picked at the L4 stage, and transferred every 8 hours to a new plate, until sperm exhaustion. 24 hours after transfer, unhatched eggs were counted, and a further 24 hours later, larvae were counted and abnormal phenotypes recorded. After they had finished laying, F1 hermaphrodites were genotyped at the *II-R* locus, to differentiate self from cross-progeny. Self progeny provided an internal control.

RESULTS

We first outline the sampling structure of the natural *C. elegans* populations that we

followed, describing their habitat, developmental stage and density fluctuations. We then turn to the microsatellite genotypes of isolated individuals, first analyzing the molecular diversity and heterozygote frequency in each locality, and the spatial structure at different scales. We finally analyze the temporal dynamics of these populations and the dynamics of linkage disequilibrium.

Habitat and population density

Samplings: We sampled *C. elegans* in different locations in France and Portugal, including those of our previous study (BARRIÈRE and FÉLIX 2005) (Figure 1A). In several instances, samples were collected at different points within a location to probe for population structure at a small spatial scale (Figure 1B-D). In addition to compost heaps, we sampled rotting fruits in the same gardens and in the Botanical Garden in Lisbon, Portugal.

The localities of our previous study had been sampled either in 2002 or in 2004 (BARRIÈRE and FÉLIX 2005). Among those, the Franconville, Le Perreux, Le Blanc and Hermanville compost heaps were newly sampled at frequent intervals until January 2006. Only the two former yielded *C. elegans* every time (Franconville) or at most timepoints (Le Perreux). For the two latter, we could only find *C. elegans* again once or twice in 2005. We also resampled in 2005 the other localities of our previous study (Merlet, Primel, Sainte-Barbe). Each sample was named using the first letters of the location followed by the month and year of sampling (e.g. Fra-1102 for the November 2002 sampling of Franconville) (Tables 1 and S1). Given the very high selfing rate, a population is difficult to define for *C. elegans*. For the sake of simplicity, we herein refer to all *C. elegans* individuals from a single location (e.g. a single compost pile) as a population.

Habitat and developmental stage: Our sampling procedure allowed us to isolate all

C. elegans individuals from a sample, most of them within a few hours (BARRIÈRE and FÉLIX 2006). In addition to previously described habitats (compost heaps, soil and snails), we found *C. elegans* in rotting fruits fallen below their trees: apples, figs, tomatoes, plums, pears, fruits from a *Ficus* tree and from an unidentified tree in the Lisbon Botanical Garden. Stages other than dauer were found mainly in fresh compost and decaying fruits, in samples Bla-1105, HerF-1005, Lis12-0705 (Table 1).

Density at a given time: Overall, density was higher in fresh compost and decomposing fruits than in older compost or soil. The highest density was found in fresh compost in Pri-0805 (21 individuals/gram; Table S1). Within the compost heaps where several samples were taken at the same time, strong variations in *C. elegans* concentration could sometimes be found at a small scale (Table S1, Figure S1).

Density over time: Given the large variance observed between samples collected on the same day, it may be hazardous to draw a strong conclusion on temporal dynamics using timepoints when a single sample was collected. However, in Franconville and Le Perreux, where samples were taken most regularly, densities seemed to decrease consistently during winter and spring, when food was presumably scarce and temperatures low (Figure S1). Density was consistently lower in Le Perreux than in Franconville and we could not find any *C. elegans* in Le Perreux in April 2005 (Table 1). In addition, at most timepoints we could not find any *C. elegans* in Le Blanc and Hermanville compost heaps, even after processing copious amounts of compost sampled at different points in the heap (however, on several occasions we found *Caenorhabditis briggsae*, a relative of *C. elegans*).

Molecular diversity and outcrossing rate

A random subset of the isolated animals were genotyped at six microsatellite loci, named *II-R, II-L, III-R, IV-L, V-L, X-R* according to their chromosomal arm position (see Methods, HABER *et al.* 2005). Haplotypes are identified by the location name followed by a letter (Figures 4 and S2).

Molecular diversity: Diversity, as measured either by pairwise differences in microsatellite haplotype or by gene diversity over the six loci, showed striking variations between samples (Table 2). Whereas some samples (Obe-1005, Lis12-0705, all Fra, all HerC) were very polymorphic, others (Per-0205, Per-0604, Per-0905, Mer1-0902, Mer2-0902, Mer3-0902, HerF-1105) were monomorphic. The low-density compost heap in Le Perreux was found to be less polymorphic than high-density compost heap in Franconville.

Heterozygote frequency and outcrossing rate: Out of the 540 individuals that we genotyped without prior selfing (see Methods), we found 10 heterozygotes, five of which were found in the Bla-1105 sample (Table 2) (this does not include the already described heterozygotes in samples Per-1004 and Fra-1004; BARRIÈRE and FÉLIX 2005). We calculated the equilibrium inbreeding coefficient f, an estimator of the inbreeding coefficient F_{IS}, from which we deduced the selfing rate s and outcrossing rate (1-s) (see Methods). Estimated outcrossing rates ranged from 0% to 7.6% (Bla-1105) (Table 2). Calculated over all diploid genotypes, the mean outcrossing rate of the different populations was 1.7% (c.i.: 1.1-2.5%). This is very consistent with our previous results (1.3%) (BARRIÈRE and FÉLIX 2005). In addition, the present results suggest that outcrossing rates vary between populations and provide an example of a population (Bla-1105) where outcrossing rates are significantly higher than in other samples, and closer to those found in SIVASUNDAR and HEY (2005) (*ca* 20%).

Male frequency and genotype: We found two males out of 993 individuals (samples Fra-0805 and Obe-1005). Together with our previous samplings (BARRIÈRE and FÉLIX 2005), we obtained a total of four males for 2269 individuals, yielding an overall male frequency of 0.18%.

We placed each of the two males with *unc-119* mutant hermaphrodites (bearing a recessive mutation resulting in uncoordinated movements) and scored non-Uncoordinated progeny. The male from the Fra-0805 sample sired no progeny, while the cross with the male from the Obe-1005 sample succeeded. Four of its progeny were genotyped at six loci and were found to be identical, indicating that the male was a homozygote at these six loci. The male may thus have been either a spontaneous male resulting from X chromosome non-disjunction, or a male resulting from biparental inbreeding (mating among identical genotypes).

Structure at different spatial scales

Within a single compost heap: In order to know whether the compost piles that we followed most closely (Franconville and Le Perreux) could each be considered homogenous, we measured spatial differentiation in genotype between samples from six different points within the pile (separated by 10 to 50 centimeters). In the three cases when enough individuals were isolated from the different samples and showed polymorphism, no significant genetic structure was found (Table 3). Thus, even though there were strong differences in density within a compost heap, we found no evidence of genetic structure within a heap. We therefore considered nematodes from a given compost pile to be genetically homogenous (what we call a population).

When comparing the genotypes of *C. elegans* individuals found in soil in Merlet or compost in Franconville, Primel and Sainte-Barbe, to those found on invertebrates (snails or isopods) sampled on the same habitat, no significant genotypic substructure was found either (Table 3).

Within a garden: structure at the scale of tens of meters: Strong and highly significant genetic structure at a given timepoint was found at the next spatial scale, when comparing samples within the same garden, in Hermanville (10-20 m), Merlet (10-100m), Lisbon (100-300 m) and between the Primel and Sainte-Barbe compost heaps (1 km) (Table 3). This was in agreement with the very strong structure observed within the Merlet location in 2002 (BARRIÈRE and FÉLIX 2005).

Large-scale structure and haplotype sharing: At the global scale (100-1000 km scale), differentiation was significant between locations (Table 3), confirming our previous results (F_{ST}=0.78, BARRIÈRE and FÉLIX 2005). Except for two neighboring Primel and Sainte-Barbe locations (1 km apart), which share one haplotype at different timepoints (see below), only three instances of haplotype-sharing between locations were found (Bla-B = Her-A, Bla-N = Her-K, Bla-I = Mer-E).

In summary, *C. elegans* appeared to show no structure at a very small scale of a few centimeters, and very strong structure at scales above 10 meters; at a larger scale of hundreds of kilometers, structure appears weaker again, as previously observed (BARRIÈRE and FÉLIX 2005; CUTTER 2006).

Temporal dynamics of genotypes

The main aim of our sampling was to analyze population dynamics over time. Genetic

diversity fluctuated over time for a given locality, especially in Le Perreux, where it reached zero at several timepoints (Table 2). Most strikingly, allele frequencies varied dramatically, sometimes even between samples collected only six weeks apart. For the two localities followed most closely, Franconville and Le Perreux, we conducted pairwise comparisons of multilocus genotypes of consecutive samples. In Franconville, a single pair of consecutive samples was significantly differentiated, whereas in Le Perreux, several pairs were significantly differentiated (asterisks in Figures 2B and 3B; Table 4).

In Franconville, where *C. elegans* could be isolated at all timepoints, the same major alleles at the three polymorphic loci (*II-R, II-L* and *III-R*) were conserved over time (Figure 2A). Allele frequencies fluctuated, but overall the population was rather stable.

By comparison, Le Perreux showed much stronger variations; after the density decreased in late winter 2004, and sample Per-0405 yielded no *C. elegans*, new alleles were found at loci *II-R* and *X-R* in June 2005 (Figure 3A). In July 2005, another replacement took place (at loci *X-R* and *V-L*). In September, the alleles and haplotypes found before April returned. Alleles found in June 2005 were found again in a single individual in December 2005 (Figure S2, haplotype Per-G). This strongly suggested an extinction of the population, followed by recolonization by new genotypes.

For the five other localities, we analyzed samples separated by one to three years. Merlet 1 showed no significant differentiation over three years, the same major haplotype Mer-E being present in both samples (Table 4, Figure S2D). Hermanville showed a marginally nonsignificant differentiation after Bonferroni correction, with a single minor haplotype (Her-D) being shared between the two timepoints (Table 4, Figure S2B). The three other localities (Le Blanc, Primel and Sainte-Barbe) showed very significant temporal differentiation (Table 4). Le Blanc witnessed an increase in genetic diversity that suggested an input from migration (several new alleles at several loci, Figure 4): the mutation rate was

measured for locus *II-L* at 1.8 x 10^{-4} and for locus *IV-L* at 2.7 x 10^{-4} per generation (FRISSE 1999) and these rates are too low to account for the increase in diversity observed in this locality. The Primel/Sainte-Barbe sampling locations (1 kilometer apart) each showed strong temporal differentiation. As noted earlier, the major haplotype (PriBar-B, Figure 4) of Pri-1004 was found at a high frequency in sample Bar-0805 while it was absent in samples Pri-0805 and Bar-1004, an indication that this temporal differentiation was in part due to migratory input.

Thus, natural populations of *C. elegans* can display dramatic changes in allele frequencies over short periods of time. In at least two localities (Le Perreux, Le Blanc), these variations were associated with density decline and subsequent recolonization events. In contrast, in two other locations with larger *C. elegans* populations (Franconville, Merlet 1) stable genotypes were maintained over several years.

Strikingly, alleles at different loci remained associated over time within a locality, suggesting little effective outcrossing (Figure 2B and S2). We therefore investigated the dynamics of linkage disequilibrium between loci.

Linkage disequilibrium

Multilocus linkage disequilibrium within a sample: The level of linkage disequilibrium between all loci was very high and significant for all polymorphic samples except Fra-1204 (Table S3). This is consistent with our previous observations based on AFLP data (BARRIÈRE and FÉLIX 2005).

Linkage disequilibrium over time: A striking fact was the maintenance of very strong linkage disequilibrium between the same alleles for loci on different chromosomes

over three years in Franconville. Indeed, the same two major multilocus genotypes in linkage disequilibrium were found throughout the three years, with very few recombinant genotypes (Figure 2B).

We looked for any evidence of decay of linkage disequilibrium in this locality. We calculated the linkage disequilibrium D between loci *II-R* and *III-R*, the two biallelic loci, and the associated D' (scaled by the maximum linkage disequilibrium level possible within the sample). We chose to work with the classical coefficient of linkage disequilibrium D because it is a simple statistic, whose decay equation is trivial (see below). Linkage disequilibrium levels were indeed very high, and remained high for more than three years (Figure 5). The sign of D' was the same in all samples, indicating that the polarity of linkage disequilibrium was conserved. The D' measure for the Fra-1204 sample was significantly lower than for later samples (comparing Fra-1204 with Fra-0106, the p-value=1.69×10^{-5}), which would indicate an increase (not a decrease!) in linkage disequilibrium over time. If we considered sample Fra-1204 as an anomaly and discarded it, the confidence intervals of D' for all samples were compatible. In any case, high linkage disequilibrium between the same alleles was maintained over three years. Given the density observed (see Table 1), the census size at the scale of the compost pile must be in the order of tens of thousands; therefore, drift alone cannot explain the absence of increase in recombinants frequency.

The outcrossing rate in Franconville was typical of results from all populations (0.9% over all timepoints), so the maintenance of linkage between alleles over such a long period of time was particularly puzzling. To test whether selfing alone could explain this level of linkage disequilibrium, we calculated the number of generations that would be compatible with the estimated outcrossing rate and the maintenance of linkage disequilibrium. The maximum decay of linkage disequilibrium compatible with our data would be from D'_0=1 in generation 0 (upper bound of confidence interval for sample Fra-1102) down to D'_N=0.947 in

generation N (lower bound for sample Fra-0106). Using these numbers, we estimated (see Methods) the maximum number of generations N_{max} as 6.0 generations over 38 months (one every 6.3 months). With the average outcrossing rate over all our samples, N_{max} would be 4.2 generations (one every 9 months). These values are hardly compatible with the known generation time and life expectancy in *C. elegans*, which in standard laboratory conditions are 3.5 days and two weeks, respectively. The generation time could be much longer in the wild, depending on temperature, food availability, occurrence of diapause, etc. However, it appears unlikely that *C. elegans* reproduces with an average of two generations per year. Therefore, the maintenance of linkage between loci must be explained by other factors.

Selection against heterozygotes and recombinants? Outcrossing rates could be variable over time, for example seasonal or with a positive correlation between outcrossing and density. However, this cannot explain an increase in linkage disequilibrium. One possible alternative explanation would be partial reproductive isolation between the two haplotypes, either prezygotic (a lower rate of mating) or postzygotic (lower fitness of progeny from a cross between the two haplotypes). The former possibility was contradicted by the fact that recombinants between the two haplotypes were found, including some heterozygotes. To test the hypothesis of postzygotic isolation, we crossed two strains representing the two major haplotypes at the first timepoint (JU360 and JU361) and compared self- versus cross-progeny for progeny number and survival in the F2 generation. Embryonic lethality was high (4%) in the F2 progeny of F1 self- and cross-progeny; however, no significant difference was found between self- and cross-progeny in terms of brood size (means: 191.0 and 205.7, respectively, p-value: 0.42), embryonic lethality nor other obvious defects (Table S4).

DISCUSSION

The present results confirm previous findings on *C. elegans* genetic diversity, geographical structure and outcrossing rate based on heterozygote frequency. Most interestingly, they provide evidence for 1) highly dynamic populations, sometimes undergoing extinction and recolonization events, and 2) maintenance of linkage disequilibrium between loci over several years. We discuss the possible consequences of this dynamic aspect of *C. elegans* populations for its genetic and phenotypic evolution.

Spatio-temporal structure of *C. elegans* populations: metapopulation dynamics

The systematic sampling of several locations at different timepoints allowed us to obtain a dynamic picture of *C. elegans* natural populations. The metapopulation dynamics that we observed involve several temporal and spatial scales. The temporal scale of population turnover appears to be a few weeks or months, consistent with *C. elegans* being found on ephemeral habitats, such as fruits rotting below their tree or decaying snails, or at a longer timescale compost heaps and the surroundings of trees during the ripening period. The spatial scale of founding individual migration appears to be above one meter, with significant migration over very long distances, and no correlation between genotypic divergence and geographic distance (BARRIÈRE and FÉLIX 2005; CUTTER 2006). High selfing rates may increase apparent levels of structure by reducing the effective number of migrants (NORDBORG 1997).

Different migration modes may operate at different spatial scales. Nematodes in the soil are able to move by themselves over a mean distance of 15 centimeters, and sometimes one meter, in a month (ROBINSON 2004), which could explain the lack of structure at small scale. At a larger scale, migration may occur through vectors such as invertebrate associates

(snails, isopods, etc.; KIONTKE and SUDHAUS 2006); the sample HerD-1105, recovered from a fly bait (see Table 1), supports this idea. Wind was described to be a potential long-distance (hundreds of meters to a few kilometers) migration vector for plant parasitic nematodes (WHITE 1953), and dust storms can be responsible for considerable movement of dauer larvae. Indeed, dauers of *C. elegans* can survive dessication for several days at room temperature (KIONTKE and SUDHAUS 2006) and may use migration vectors that would appear unfit at first glance. Since we found a high density of *C. elegans* in rotting fruits, fruits may also be an efficient way to migrate – with the help of humans, flies or birds – over long distances. Overall, the association of *C. elegans* with human-related habitats indicates that human activities could be responsible for large scale migration. The fact that *C. elegans* was seldom found in soil supports the idea of a patchy distribution of the species, with migration being a critical survival factor.

Outcrossing and maintenance of strong linkage disequilibrium

Three different measures provide information on the relative occurrence of selfing and outcrossing: heterozygote frequency, male frequency and linkage disequilibrium. Heterozygote and male frequencies provide a measure at the short temporal scale of the previous generations. Our new estimates of heterozygote frequencies, yielding a global estimate of 1.7% outcrossing (c.i.: 1.1-2.5%), confirm our previous ones (BARRIÈRE and FÉLIX 2005). It is improbable that many of these heterozygotes are due to mutational events, because half of them are heterozygotes at several loci (Dataset S1). If we consider selfing to be constant in the species, this 1.7% outcrossing rate predicts a male frequency of 0.85% (c.i.: 0.55-1.25%), not counting spontaneous males, whereas the observed male frequency is only 0.18% (95% c.i.: 0.05-0.45%). We cannot completely rule out that some males were missed in our sampling procedure, but the large discrepancy suggests that outcrossing rates vary over

time and between different populations, as also suggested by the variation in outcrossing rate estimates among populations (Table 2).

Much more divergent is the 100-fold lower outcrossing rate estimate (10^{-4}) based on static estimates of linkage disequilibrium between loci in a local population (BARRIÈRE and FÉLIX 2005) or among worldwide isolates (CUTTER 2006). We find a very strong linkage disequilibrium between loci located on different chromosomes for most samples (Table S3). CUTTER (2006) found similarly strong linkage on a worldwide scale (multilocus linkage disequilibrium I_A^S=0.29). Within a population of constant size, significant nonrandom associations between loci can appear by mutation and drift. In rapidly growing populations, like those undergoing metapopulation dynamics, such nonrandom associations are expected to be rare (SLATKIN 1994). Therefore, the observed linkage disequilibrium in *C. elegans* must have been present since the foundation of the population, possibly after colonization by two genotypes. Such a high level of linkage disequilibrium would then be expected to decay over time as a function of the outcrossing rate.

In our temporal surveys, we found no evidence of linkage disequilibrium decay over three years in the Franconville population. These observations are not compatible with the outcrossing rate estimated from heterozygote frequencies, unless the generation time is more than half a year. We found *C. elegans* mostly in the dauer stage, which could be responsible for a huge increase in generation time (in laboratory conditions, dauers may live up to eight months; C. Braendle, personal communication). However, the occurrence of only six generations in 38 months is improbable.

Several mechanisms can explain the discrepancy between the short-term outcrossing rate measured by heterozygote frequency and the maintenance of high linkage disequilibrium. Population structure may explain high linkage disequilibrium among, but not within populations. The sampled populations could however be sink populations, receiving a

constant flow of migrants from two populations, each monomorphic for one major haplotype: a Wahlund effect (WAHLUND 1928) could then explain the absence of decay of linkage disequilibrium at a given timepoint. Linkage should ultimately decay over time in this population, given that some sampled individuals appeared to be part of the reproductive pool (non-dauer stages). Our observations may however reflect stochastic effects associated with a small effective population size. Finally, an alternative hypothesis to explain our observations is selection acting against the cross-progeny (F1 or later generations). In laboratory conditions, we failed to find evidence of a strong effect on brood size of recombination of the two major Franconville haplotypes, yet it is possible that another character affecting fitness in natural conditions may be affected. Outbreeding depression was indeed observed between *C. elegans* isolates, including within some of our local sets (DOLGIN et al., 2007). Thus, a possible explanation for maintenance of linkage disequilibrium is selection against heterozygotes or recombinants.

By comparison, in *Drosophila melanogaster*, linkage disequilibrium is very low and decays within 1 kb (LONG *et al.* 1998). Even in a highly selfing species like *Arabidopsis thaliana*, linkage disequilibrium is lower than in *C. elegans*, at least at a large geographical scale: indeed, at this global scale, linkage disequilibrium is undetectable between different chromosomes and decays within approximately 50-250 kb for linked loci (NORDBORG *et al.* 2002, 2005); linkage disequilibrium on a short region of 170 kb is I_A^S=0.179 (HAUBOLD *et al.* 2002), weaker than on the complete genome of *C. elegans,* and unlike in *C. elegans* (CUTTER 2006), some of it may be the result of spatial population structure (SCHMID *et al.* 2006). At a small scale within a patch of *A. thaliana*, linkage disequilibrium appears extensive, as in *C. elegans* (BERGELSON *et al.* 1998; NORDBORG *et al.* 2002; STENOIEN *et al.* 2005).

Possible consequences of *C. elegans* population dynamics on its phenotypic evolution

Both demographic and genetic results allow us to infer several consequences for the

genetic and phenotypic evolution of *C. elegans*. The low outcrossing rate implies that alleles mostly occur in a homozygous state, hence purging of strongly deleterious recessive mutations should occur more readily than in outcrossing populations. Populations experiencing bottlenecks are likely to fix slightly deleterious mutations by genetic drift, but the strong reexpansion regimes that follow may allow compensatory mutations to occur. Recent experimental evolution studies in *C. elegans* revealed that significant increase in fitness could already be seen after 10 generations (ESTES and LYNCH 2003). A transient loss of fitness, or of robustness of a given phenotypic character, followed by compensatory evolution may be a frequent mechanism of exploration of the genotype-phenotype landscape in *C. elegans*. Furthermore, the almost exclusive selfing of *C. elegans* would be expected to favor co-evolution of its entire genome, thus resulting in outbreeding depression when outcrossing actually occurs (AGRAWAL 2006).

In addition, if the *C. elegans* metapopulation comprises source and sink populations, adaptation in the sink populations, which are doomed to extinction, is not relevant to future generations of the species as a whole, which would adapt only to source environments. Identification of sink and source environments is thus crucial for the study of adaptive traits of *C. elegans*.

The very low overall genetic diversity of *C. elegans* (BARRIÈRE and FÉLIX 2005; HABER *et al.* 2005; SIVASUNDAR and HEY 2005; CUTTER 2006) cannot be explained by the mere two-fold reduction due to selfing; however, the observed metapopulation dynamics, in association with high rates of selfing, may result in selective sweeps that affect the whole genome and thus drastically reduce overall genetic and phenotypic diversity (CHARLESWORTH and CHARLESWORTH 1998). The population dynamics of *C. elegans* will thus affect molecular evolution patterns by reducing genetic diversity, increasing linkage disequilibrium, and potentially allowing the fixation of slightly deleterious mutations, which may then be

compensated at the same or another locus.

Acknowledgments

We are very grateful to our colleagues and frequent compost providers, J.-A. Lepesant and C. Pieau, and to all those who maintained the gardens that were sampled. We thank D. Higuet and B. Toupance for help and advice with the analyses, and C. Braendle, D. Charlesworth, A. Cutter, E. Dolgin as well as two reviewers for helpful comments on the manuscript. A. B. was supported by the Ministry of Research of France and the Association pour la Recherche sur le Cancer. This work was supported by the Centre National de la Recherche Scientifique and the Ministry of Research of France through a Biological Resource Center grant.

Table 1. Sampling of *C. elegans* populations.

Summary description of *C. elegans* population sampling. 'Location': name of sampling location; 'Sample': name of sample; 'Date': date of sampling; 'Mass': mass of sample (in grams); 'Ind': number of *C. elegans* individuals recovered, 'Dens': density of sample (individuals per gram); 'Nature': nature of sample and habitat; 'Stages/remarks': occurrence of different developmental stages when we were able to determine it, and remarks about the sample. The four larval stages are indicated as L1-L4. 'd': dauer. 'ad': adult. See Table S1 for more details, especially of subsamples.

Location	Sample	Date	Mass	Ind	Dens	Nature	Stages
Sainte Barbe	Bar-1004	3 Oct 04	15	98	6.53	compost	described in BARRIÈRE and FÉLIX (2005)
	Bar-0805	27 Aug 05	27.3	38	1.39	compost	2 L4
		27 Aug 05		4		isopods	3 d
Primel	Pri-1004	3 Oct 04	15	32	2.13	compost and isopods	described in BARRIÈRE and FÉLIX (2005)
	Pri-0805	27 Aug 05	52	161	3.09	compost	11 d, 2 L3, 1 L4, 1 ad
		27 Aug 05		7		isopods	6 d
		27 Aug 05		6		*Helix aspersa* snail	
Le Blanc	Bla-0802	25 Aug 02		13		compost	described in BARRIÈRE and FÉLIX (2005)
		28 Mar 05	35	6	0.17	compost	3 L3
		12 Jun 05	115	1		compost	
		15 Aug 05	131.5	0		compost	
	Bla-1105	1 Nov 05	98	23	0.23	compost	3d, 3 L3, 8 L4, 9 ad
Hermanville	HerC-0902	22 Sep 02		12		compost	described in BARRIÈRE and FÉLIX (2005)
	HerC-1105	8 Nov 05	20	38	1.9	compost - mixed L3-L4-adults	
	HerF-1005	1 Nov 05		47		figs	28 d, 6 L4, 10 ad
	HerF-1105	8 Nov 05		12		figs	high density
	HerD-1105	8 Nov 05		1		recovered from a cleaned fig placed as a fly bait above compost	
	HerS-0603	1 Jun 03		1		caught in soil, near a snail	
Franconville	Fra-1102	16 Sep 02		12		compost	described in BARRIÈRE and FÉLIX (2005)
	Fra-1004	6 Oct 04	13	130	10	compost	described in BARRIÈRE and FÉLIX (2005)
	FraS-1004	6 Oct 04		4		snails next to compost	described in BARRIÈRE and FÉLIX (2005)
	Fra-1204	8 Dec 04	9.5	43	4.53	compost	some non-dauer
	Fra-0205	14 Feb 05	31	40	1.29	compost	33 d
	Fra-0405	6 Apr 05	34	11	0.32	compost	1 L3, not all picked
	Fra-0505	20 May 05	42	14	0.33	compost	
	Fra-0605	7 Jun 05	42	184	4.38	compost	
	Fra-0705	19 Jul 05	59	64	1.08	compost	55 d
	Fra-0805	1 Aug 05	128	86		compost	all dauer, 1 male
	Fra-0905	23 Sep 05	125	29		compost	19 d, 1 L3, 7 L4, 1 ad
	Fra-0106	7 Jan 06	99	57	0.57	compost	43 d
Lisbon	Lis8-0705	10 Jul 05		45		dates and leaf litter	1 L2, 38 d, 4 L3, 2 L4
	Lis12-0705	10 Jul 05		77		figs from *Ficus isophlebia* and	3 L1, 8 L2, 55 d, 1 L3, 1 L4, 8 ad

						associated litter	
	Lis14-0705	10 Jul 05		17		compost	4d, 2 L4, 1 ad
Merlet	Mer1-0902	8 Sep 02		13		soil below hackberry tree	described in BARRIÈRE and FÉLIX (2005)
	Mer2-0902	8 Sep 02		4		snails on mulberry tree	described in BARRIÈRE and FÉLIX (2005)
	Mer3-0902	8 Sep 02		4		compost	described in BARRIÈRE and FÉLIX (2005)
	Mer1-1005	23 Oct 05	49	12	0.24	soil below hackberry tree	10 d
		6 Oct 04	-	9		*Pomatias* snails	
	Mer3-1005	23 Oct 05	29	1	0.03	compost	
	Mer4-1005	23 Oct 05		1		figs	
	Mer5-1005	23 Oct 05		9		apples	1 L2, 1 ad
Obernai	Obe-1005	3 Oct 05		29		fruit in orchard	1 L1, 8 d, 4 L4, 4 ad
Le Perreux	Per-0704	7 Jul 04		6		compost	described in BARRIÈRE and FÉLIX (2005)
	Per-1004	5 Oct 04	17	36	2.12	compost	described in BARRIÈRE and FÉLIX (2005)
	Per-1204	14 Dec 04	30	10	0.33	compost	1 L4, 1 ad
	Per-0205	7 Feb 05	79	13	0.16	compost	9 d
	Per-0405	5 Apr 05	62	0		compost	
	Per-0605	8 Jun 05	68	6	0.09	compost	6 d
	Per-0705	13 Jul 05	20	10	0.5	compost	
	Per-0905	26 Nov 05	102	45	0.44	compost	all dauer
	Per-101205	10 Dec 05	79	2	0.03	compost	2 d
	Per-1205	19 Dec 05	80	13	0.16	compost	13 d

Table 2. Molecular diversity and outcrossing rates in the sampled *C. elegans* populations.

'Inbred?': whether individual genotypes were scored after laboratory inbreeding ('i', inbred strains) or without ('h', potentially heterozygous). 'N': number of genotyped individuals. 'Pair.diff.': mean number of different pairwise microsatellite loci; 'H': gene diversity (calculated with Arlequin; EXCOFFIER *et al.* 2005) and its standard deviation ('s.d.'); 'Θ_{hom}': theta parameter calculated after homozygosity and its standard deviation; 'Het': number of heterozygotes; 'Ho': observed heterozygosity; *f*: equilibrium inbreeding coefficient (estimator of *Fis* after WEIR and COCKERHAM, 1984, calculated with GDA**;** LEWIS and ZAYKIN, 2001) and its 95% confidence interval ('c.i.'); *s*: selfing rate calculated after *f* and its 95% confidence interval.

			Genetic diversity					Outcrossing					
Sample	Inbred?	N	Pair.diff.	H	s.d.	Θ_{hom}	s.d.	Het	Ho	*f*	c.i.	*s*	c.i.
Bar-1004	h	15	1.251	0.699	0.084	1.774	0.739	0	0.000	1.000		1.000	
Bar-0805	h	14	2.152	0.605	0.052	1.151	0.257	1	0.067	0.940	(0.877-1)	0.969	(0.934-1)
Bla-0802	i	13	0.282	0.271	0.099	0.277	0.138						
Bla-1105	h	24	3.437	0.832	0.046	4.014	1.432	5	0.208	0.858	(0.819-0.901)	0.924	(0.901-0.948)
Fra-1102	i	12	1.621	0.507	0.093	0.769	0.289						
Fra-1004	h	12	1.522	0.754	0.058	2.377	0.793	0	0.000	1.000		1.000	
Fra-1204	i	12	1.621	0.797	0.067	3.110	1.394						
Fra-0205	i	12	1.712	0.609	0.087	1.171	0.438						
Fra-0405	h	6	1.818	0.727	0.109	2.054	1.190	0	0.000	1.000		1.000	
Fra-0505	h	12	1.779	0.757	0.060	2.427	0.846	2	0.167	0.865	(0.755-1)	0.927	(0.86-1)
Fra-0605	h	82	1.003	0.533	0.045	0.852	0.156	1	0.012	0.988	(0.969-1)	0.994	(0.984-1)
Fra-0705	i	12	1.152	0.522	0.099	0.815	0.327						
Fra-0905	h	18	1.594	0.679	0.024	1.613	0.185	0	0.000	1.000		1.000	
Fra-0106	h	57	0.896	0.504	0.051	0.760	0.158	0	0.000	1.000		1.000	
HerC-0902	i	12	3.758	0.812	0.040	3.436	0.979						
HerC-1105	h	19	2.504	0.762	0.068	2.502	1.002	0	0.000	1.000		1.000	
HerF-1005	h	45	0.261	0.170	0.053	0.153	0.058	0	0.000	1.000		1.000	
HerF-1105	h	11		0.000				0	0.000				
Lis8-0705	h	9	1.098	0.680	0.108	1.616	0.839	0	0.000	1.000			
Lis12-0705	h	20	3.103	0.836	0.042	4.125	1.386	0	0.000	1.000			
Lis14-0705	h	9	0.993	0.392	0.133	0.480	0.268	0	0.000	1.000		1.000	
Mer1-0902	i	11		0.000									
Mer1-1005	h	18	0.324	0.210	0.088	0.198	0.104	0	0.000	1.000		1.000	
Mer2-0902	i	4		0.000									
Mer3-0902	i	4		0.000									
Mer3-1005	h	1		0.000				0	0.000				
Mer4-1005	h	1		0.000				0	0.000				
Mer5-1005	h	4	2.714	0.857	0.082	4.935	3.621	0	0.000	1.000		1.000	
Obe-1005	h	18	0.979	0.884	0.034	6.428	2.350	1	0.056	0.973	(0.92-1)	0.986	(0.958-1)
Per-0604	h	10		0.000				0	0.000	1.000		1.000	
Per-1004	h	12	0.464	0.304	0.115	0.326	0.176	0	0.000	1.000		1.000	
Per-1204	i	9	2.028	0.758	0.077	2.440	1.100						
Per-0205	i	12		0.000									

Per-0605	h	6	0.303	0.303	0.148	0.324	0.226	0	0.000	1.000	1.000
Per-0705	i	10	0.556	0.779	0.061	2.766	1.056				
Per-0905	h	44		0.000				0	0.000		
Per-1205	h	13	1.329	0.394	0.107	0.484	0.218	0	0.000	1.000	1.000
Pri-1004	h	35	1.559	0.601	0.069	1.134	0.335	0	0.000	1.000	1.000
Pri-0805	h	11	2.615	0.537	0.090	0.867	0.320	0	0.000	1.000	1.000

Table 3. Spatial structure of genetic differentiation at different scales.

'p-value': p-value of differentiation test after Bonferroni correction. Values below 0.05 indicate a significant spatial structure. 'θ': estimator of F_{St} after WEIR and COCKERHAM (1984) with its confidence interval ('c.i.').

Sample	p-value	θ	c.i.
Structure within one compost pile			
Fra-0106	0.990		
Fra-0905	0.650		
Per-1205	0.792		
Mer1-1005 (snails-compost)	0.792		
Pri-0805 (snails-compost)	0.792		
Fra-1004(snails-compost)	0.133		
Bar-0805 (isopods and compost)	0.420		
Within one garden			
Lisbon	<0.001	0.366	0.189-0.504
Hermanville	<0.001	0.255	0.207-0.295
Merlet 05	<0.001	0.776	0.598-0.886
Primel–Sainte-Barbe 04	<0.001	0.591	0.453-0.664
Primel–Sainte-Barbe 05	<0.001	0.443	0.270-0.670
Merlet 02	<0.001	1.000	1.000-1.000
Large scale			
All locations (Jul-Nov 05)	<0.001	0.532	0.445-0.605

Table 4. Temporal structure of genetic differentiation.

'p-value': p-value of differentiation test after Bonferroni correction; 'θ': estimator of F_{ST} after WEIR and COCKERHAM (1984) with its confidence interval ('c.i.').

Samples	p-value	θ	c.i.
HerC-1102–HerC-1105	0.056	0.159	0.080-0.216
Mer1-0902–Mer1-1005	1		
HerF-1005–HerF-1105	1		
Pri-1004–Pri-0805	<0.001	0.494	0.253-0.639
Bar-1004–Bar-0805	0.032	0.254	0.095-0.382
Between Fra samples			
Fra-1102–Fra-1004	0.662		
Fra-1004–Fra-1204	0.963		
Fra-1204–Fra-0205	1		
Fra-0205–Fra-0405	1		
Fra-0405–Fra-0505	1		
Fra-0505–Fra-0605	0.275		
Fra-0605–Fra-0705	<0.001	0.505	0.430-0.546
Fra-0705–Fra-0905	1		
Fra-0905–Fra-0106	0.275		
Between Per samples			
Per-0604–Per-1004	1		
Per-1004–Per-1204	0.963		
Per-1204–Per-0205	0.065		
Per-0205–Per-0605	<0.001	0.965	0.891-0.979
Per-0605–Per-0705	<0.001	0.842	0.736-0.931
Per-0705–Per-0905	<0.001	0.956	0.869-0.993
Per-0905–Per-1205	0.108		

Figure 1. Sampling locations.

A. Map of sampling locations in mainland France and Portugal. B. Sketch of the Hermanville sampling location. C. Sketch of the Lisbon sampling location. D. Sketch of the Merlet sampling location. Scales are indicated for each map.

Figure 2. Temporal survey of allele and multilocus genotype frequencies in the Franconville population.

A. Allele frequencies at loci *II-R*, *II-L* and *III-R* (the most polymorphic loci in this population) over time in the Franconville population. The repeat number is indicated for each locus on the right. The number of genotyped individuals (N) is indicated below each timepoint (horizontal axis). B. Frequencies of multilocus genotypes for the major alleles at the same three loci. Individuals showing a recombination between the major genotypes are indicated as recombinants. Rare haplotypes (below 2% when considering all timepoints) were removed from this analysis. For more detailed data, see Figure S2G.

Figure 3. Temporal survey of allele and multilocus genotype frequencies in the Le Perreux-sur-Marne population.

A. Allele frequencies at loci *II-R*, *V-L* and *X-R* in the Le Perreux population, displayed as in Figure 2. For locus *II-R* in sample Per-0705, amplification repeatedly failed for several individuals (indicated as '?'). B. Frequencies of multilocus genotypes for the major alleles at three loci in Le Perreux. No evidence of recombination between the three major genotypes was found. Rare haplotypes (below 2% when considering all timepoints) were removed. Asterisks indicate significant differentiation between consecutive samples. For more detailed data, see Figure S2F.

Figure 4. Multilocus genotypes in the Primel/Sainte-Barbe and Le Blanc populations at two timepoints.

A. Haplotype frequencies in Primel (Pri) and Sainte-Barbe (Bar) in October 2004 and August 2005. N: number of genotyped individuals. Haplotypes are identified by their alleles (number of repeat) at each locus, in the following order: *II-R, V-L, II-L, III-R, IV-L, X-R*. Each haplotype is identified by a letter code common to both locations (PriBarA-X); haplotype PriBar-B (light grey) is found in both Pri-1004 and Bar-0805 samples. B. Haplotype frequencies in Le Blanc in 2002 and 2005. Each haplotype is identified by a letter code (BlaA-Q). Haplotype Bla-B (dark grey) was found on both dates.

Figure 5. Evolution of linkage disequilibrium over time in the Franconville population.

Linkage disequilibrium *D'* was measured at different timepoints. Error bars delimit the 95% confidence interval.

Supporting information

Dataset S1. Multilocus genotype of each individual.

Multilocus genotypes of all genotyped individuals, in Genepop format, expressed as the number of microsatellite repeats at a given locus, in the order indicated at the top of the table. Heterozygotes are in bold on a red background. The name of each individual is in the first column. The 'Sample' column indicates the sample name. When several samples were taken at different points within a compost pile, the different subsamples are identified in the 'Sample substructure' column. Data from inbred strains are indicated by 2-digit alleles and data from non-inbred strains by 4-digit allele genotypes. Missing data are indicated by '00'. The first eight columns can be copied to a text file for use in Genepop or any other

compatible software. Genepop files can be converted to other formats using the online tool http://wbiomed.curtin.edu.au/genepop/genepop_op7.html.

Table S1. Samplings of *C. elegans* populations.

Detailed description of samplings of *C. elegans* populations. 'Location': name of sampling location; 'Coordinates': coordinates of sampling locations; 'Sample': name of sample; 'Date': date of sampling; 'Mass': mass of sample (in grams); 'Ind': number of *C. elegans* individuals recovered; 'Dens': density of sample (in individuals per gram); 'Nature': nature of sample and habitat; 'Stages': occurrence of different developmental stages when we were able to determine it. When several samples were taken at the same time in the same compost pile, each subsample is detailed. Samples which yielded no *C. elegans* are also detailed.

Table S2. Allele frequencies in each sample.

Frequency of alleles in all samples. N: number of individuals in sample. Alleles are sorted by locus and by repeat number. Missing data is indicated by '?'.

Table S3. Multilocus linkage disequilibrium.

Standardized index of association I_A^S (multilocus measure of linkage disequilibrium) and its significance (p-value) for all samples. The table indicates p-values after Bonferroni correction. Samples with a very low diversity (less than two polymorphic loci, less than two individuals per allele) were excluded.

Table S4. Embryonic lethality in the self- and cross-progeny of the JU360 and JU361 isolates from Franconville.

'*II-R* alleles': genotype of F1 individual at locus *II-R*, identifying self versus cross-progeny

(the 24-repeat allele is from JU361, the 28-repeat allele from JU360); 'dead emb': number of dead embryos in F2 progeny; 'adults': number of individuals reaching adulthood in the F2 progeny; 'abnormal': number of individuals showing strong abnormalities (larval lethality, severely delayed development); 'emb let': proportion of embryonic lethality and abnormal phenotypes (%).

Figure S1. Evolution of density in *C. elegans* in Franconville and Le Perreux-sur-Marne.
Density of *C. elegans* in sampling locations Franconville (A) and Le Perreux-sur-Marne (B). In some cases, several samples were taken at the same date; the density indicated is the mean, and standard error is indicated by an error bar.

Figure S2. Multilocus genotype frequencies.
Histograms of frequency of the different haplotypes in each sample, segregated by sampling location (A-G). Haplotypes are identified by their alleles at each locus, in the following order: *II-R, V-L, II-L, III-R, IV-L, X-R* and named by a letter code and a color in each location.

References

AGRAWAL, A. F., 2006 Evolution of Sex: Why Do Organisms Shuffle Their Genotypes? Current Biology **16:** R696-R704.

BARRIÈRE, A., and M.-A. FÉLIX, 2005 High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr Biol **15:** 1176-1184.

BARRIÈRE, A., and M.-A. FÉLIX, 2006 Isolation of *C. elegans* and related nematodes in *WormBook*, edited by The *C. elegans* Research Community. http://www.wormbook.org/. doi/10.1895/wormbook.1.115.1

BERGELSON, J., E. STAHL, S. DUDEK and M. KREITMAN, 1998 Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics **148:** 1311-1323.

CHARBONNEL, N., and J. PEMBERTON, 2005 A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity **95:** 377-388.

CHARLESWORTH, B., and D. CHARLESWORTH, 1998 Some evolutionary consequences of deleterious mutations. Genetica **102-103:** 3-19.

CUTTER, A. D., 2006 Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics **172:** 171-184.

DENVER, D. R., K. MORRIS and W. K. THOMAS, 2003 Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. Mol Biol Evol **20:** 393-400.

DOLGIN, E.S., CHARLESWORTH, B., BAIRD, S.E. and CUTTER, A.D., 2007 Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. Evolution, in press.

ESTES, S., and M. LYNCH, 2003 Rapid fitness recovery in mutationally degraded lines of *Caenorhabditis elegans*. Evolution **57:** 1022-1030.

EXCOFFIER, L., G. LAVAL and S. SCHNEIDER, 2005 Arlequin (version 3.0): An integrated

software package for population genetics data analysis. Evolutionary Bioinformatics Online **1:** 47-50.

FRISSE, L., 1999 Understanding the mechanisms of microsatellite formation and mutation using the model organism *Caenorhabditis elegans*, PhD thesis, University of Missouri, Kansas City.

GOUDET, J., 2001 FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat/html.

GOUDET, J., 2005 Hierfstat, a package for R to compute and test variance components and F-statistics. Molecular Ecology Notes **5:** 184-186.

GOUDET, J., M. RAYMOND, T. DE MEEUS and F. ROUSSET, 1996 Testing differentiation in diploid populations. Genetics **144:** 1933-1940.

GUILLEMAUD, T., L. MIEUZET and J. C. SIMON, 2003 Spatial and temporal genetic variability in French populations of the peach-potato aphid, *Myzus persicae*. Heredity **91:** 143-152.

HABER, M., M. SCHUNGEL, A. PUTZ, S. MULLER, B. HASERT *et al.*, 2005 Evolutionary History of *Caenorhabditis elegans* Inferred from Microsatellites: Evidence for Spatial and Temporal Genetic Differentiation and the Occurrence of Outbreeding. Mol Biol Evol **22:** 160-173.

HANSKI, I., 1999 *Metapopulation ecology*. Oxford University Press.

HAUBOLD, H., and R. R. HUDSON, 2000 LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics **16:** 847-848.

HAUBOLD, B., J. KROYMANN, A. RATZKA, T. MITCHELL-OLDS and T. WIEHE, 2002 Recombination and Gene Conversion in a 170-kb Genomic Region of *Arabidopsis thaliana*. Genetics **161:** 1269-1278.

HODGKIN, J., and T. DONIACH, 1997 Natural variation and copulatory plug formation in

Caenorhabditis elegans. Genetics **146:** 149-164.

KIONTKE, K., and W. SUDHAUS, 2006 Ecology of *Caenorhabditis* species in *WormBook*, edited by The *C. elegans* Research Community. http://www.wormbook.org/. doi/10.1895/wormbook.1.37.1

KOCH, R., H. G. VAN LUENEN, M. VAN DER HORST, K. L. THIJSSEN and R. H. PLASTERK, 2000 Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. Genome Res **10:** 1690-1696.

LEWIS, P. O., and D. ZAYKIN, 2001 Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.

LONG, A. D., R. F. LYMAN, C. H. LANGLEY and T. F. MACKAY, 1998 Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics **149:** 999-1017.

MEUNIER, C., S. HURTREZ-BOUSSES, P. DURAND, D. RONDELAUD and F. RENAUD, 2004 Small effective population sizes in a widespread selfing species, *Lymnaea truncatula* (Gastropoda: Pulmonata). Molecular Ecology **13:** 2535-2543.

NORDBORG, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501-1514.

NORDBORG, M., J. O. BOREVITZ, J. BERGELSON, C. C. BERRY, J. CHORY *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet **30:** 190-193.

NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol **3:** e196.

R DEVELOPMENT CORE TEAM, 2003 *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROBINSON, A. F., 2004 Nematode behavior and migrations through soil and host tissue, pp.

330-405 in *Nematology-Advances and Perspectives*, edited by Z. Chen, S. Chen and D. W. Dickson. CABI publishing.

SCHMID, K. J., O. TÖRJEK, H. SCHMUTHS, M. H. HOFFMANN *et al.*, 2006 Evidence for a large-scale population structure in *Arabidopsis thaliana* from genome-wid single nucleotide polymorphism markers. Theor. Appl. Genet. **112:** 1104-1114.

SIVASUNDAR, A., and J. HEY, 2005 Sampling from natural populations with RNAi reveals high outcrossing and population structure in *Caenorhabditis elegans*. Curr Biol **15:** 1598-1602.

SLATKIN, M., 1994 Linkage disequilibrium in growing and stable populations. Genetics **137:** 331-336.

STENOIEN, H. K., C. B. FENSTER, A. TONTERI and O. SAVOLAINEN, 2005 Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. Mol Ecol **14:** 137-148.

TEOTÓNIO, H., D. MANOEL and P. C. PHILLIPS, 2006 Genetic variation for outcrossing among *Caenorhabditis elegans* isolates. Evolution **60:** 1300-1305.

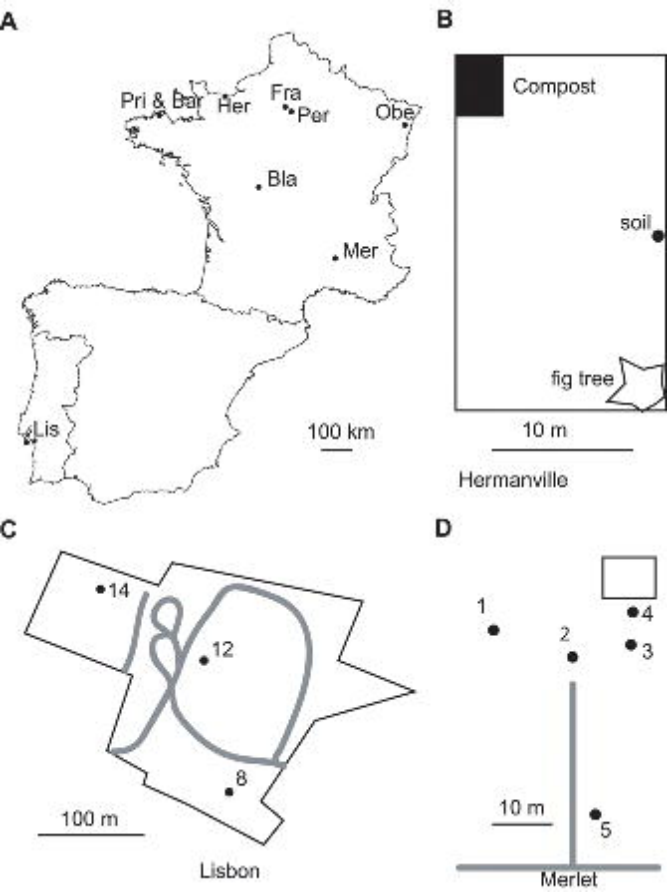
TROUVÉ, S., L. DEGEN and J. GOUDET, 2005 Ecological components and evolution of selfing in the freshwater snail *Galba truncatula*. J Evol Biol **18:** 358-370.

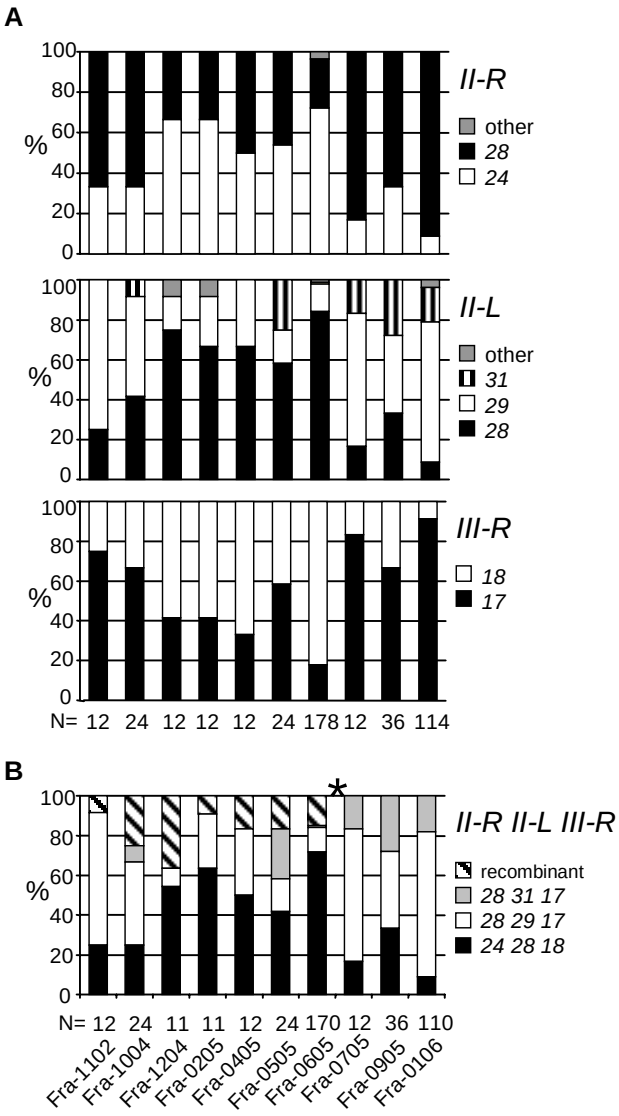
VIARD, F., F. JUSTY and P. JARNE, 1997 Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. Genetics **146:** 973-982.

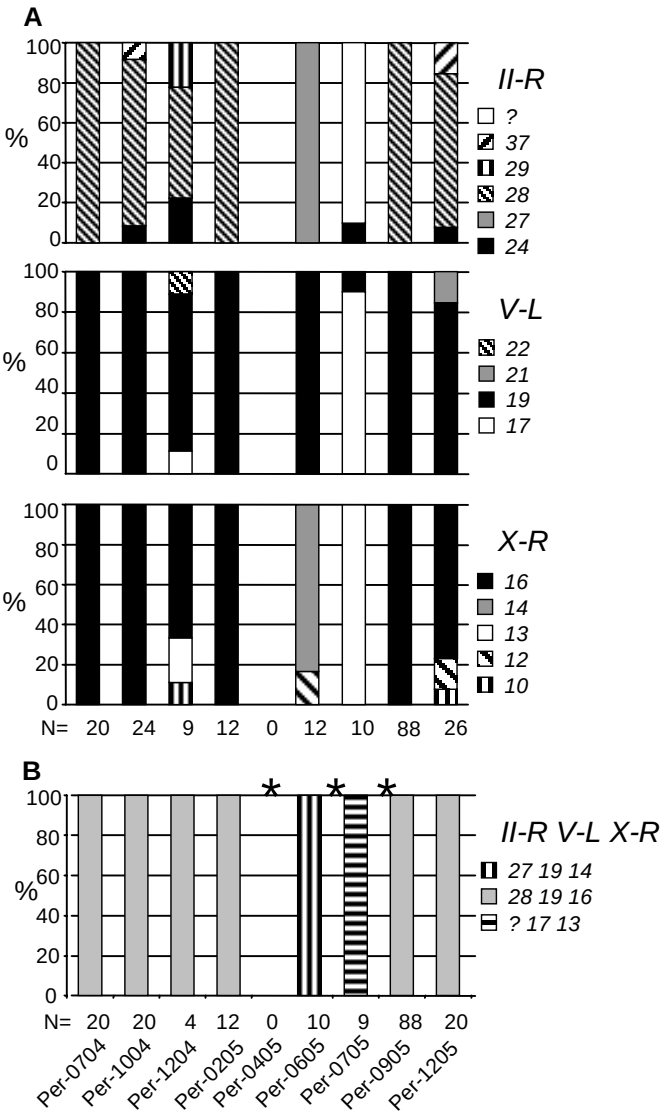
WAHLUND, S., 1928 Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **11:** 65-106.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38:** 1358-1370.

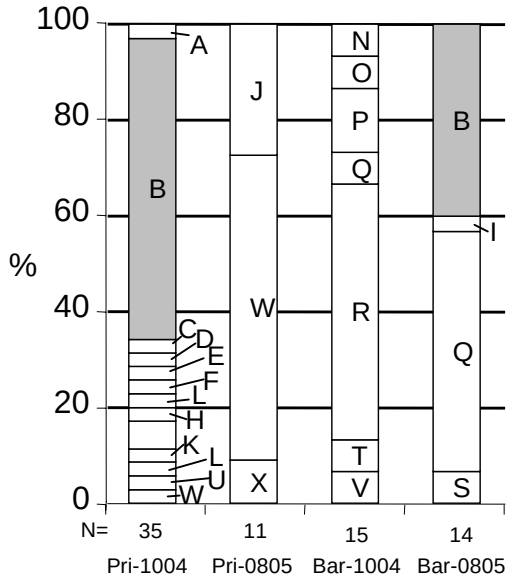
WHITE, J., 1953 Wind-borne dispersal of potato-root eelworm. Nature **172:** 686-687.


A
Pri & Bar
Her
Fra
Per
Obe
Bla
Mer
Lis
100 km
B
Compost
soil
fig tree
10 m
Hermanville
C
14
12
8
100 m
Lisbon
D
1
2
4
3
10 m
5
Merlet


A
100
80
60
40
20
0
%
II-R
other
28
24
II-L
other
31
29
28
III-R
18
17
N= 12 24 12 12 12 24 178 12 36 114
B
*
II-R II-L III-R
recombinant
28 31 17
28 29 17
24 28 18
N= 12 24 11 11 12 24 170 12 36 110
Fra-1102
Fra-1004
Fra-1204
Fra-0205
Fra-0405
Fra-0505
Fra-0605
Fra-0705
Fra-0905
Fra-0106

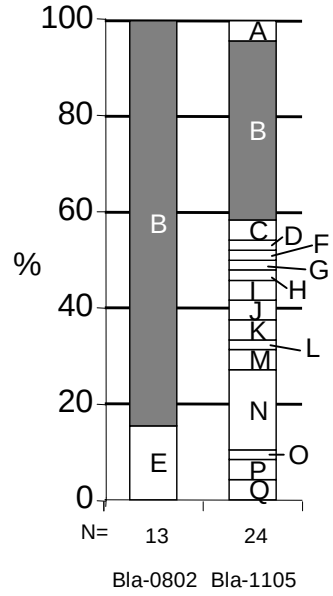

A
100
80
60
40
20
0
%
II-R
?
37
29
28
27
24
V-L
22
21
19
17
X-R
16
14
13
12
10
N= 20 24 9 12 0 12 10 88 26
B
*
*
*
II-R V-L X-R
27 19 14
28 19 16
? 17 13
N= 20 20 4 12 0 10 9 88 20
Per-0704
Per-1004
Per-1204
Per-0205
Per-0405
Per-0605
Per-0705
Per-0905
Per-1205

A

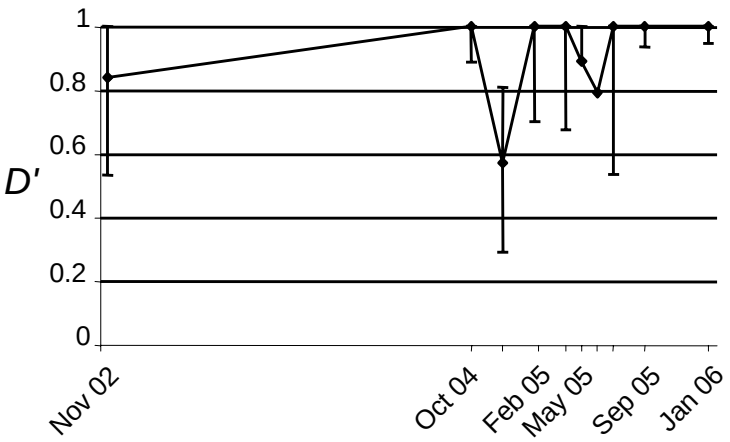


A 47 22 33 26 28 14
B 47 22 30 26 28 14
C 47 22 30 26 28 10
D 47 22 30 17 28 14
E 47 22 28 26 28 14
F 47 22 11 17 28 14
G 47 20 30 17 28 14
H 39 22 30 26 28 14
I 37 22 12 26 28 14
J 37 20 30 21 28 14
K 37 20 28 21 28 14
L 37 20 28 19 29 14
N 37 19 29 17 28 14
O 37 11 29 17 28 14
P 37 11 28 17 28 14
Q 37 11 12 21 28 14
R 37 11 12 17 28 14
S 37 11 11 17 28 14
T 28 22 30 26 28 10
U 28 19 30 17 28 14
V 28 ? 12 21 28 14
W 24 22 28 26 29 12
X 24 22 23 26 29 12

B



A 47 20 29 17 29 14
B 47 20 29 15 29 14
C 47 20 28 15 29 14
D 47 17 29 15 29 14
E 44 20 29 15 29 14
F 28 17 28 21 29 14
G 28 17 28 17 29 14
H 26 20 28 15 29 10
I 26 19 32 17 26 12
J 26 19 28 17 28 12
K 26 17 29 21 29 10
L 26 17 29 21 28 14
M 26 17 28 21 29 12
N 26 17 28 21 28 12
O 26 17 28 15 29 12
P 26 17 11 26 ? 12
Q 26 17 10 17 28 13


1
0.8
0.6
D'
0.4
0.2
0
Nov 02
Oct 04
Feb 05
May 05
Sep 05
Jan 06