Genet. Res., Camb. (2003), **82**, *pp.* 183–190. *With* 1 *figure.*
DOI: 10.1017/S0016672303006487 Printed in the United Kingdom

Coalescence times and the Meselson effect in asexual eukaryotes

ALF CEPLITIS*
Department of Conservation Biology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

(*Received 30 June 2003 and in revised form 9 September 2003*)

Summary

In asexual eukaryotes, the two allelic gene copies at a locus are expected to become highly divergent as a result of the independent accumulation of mutations in the absence of segregation. If sexual reproduction was abandoned millions of generations ago, intra-individual allelic divergences can be significantly larger than in species that reproduce sexually. Owing to the disputed existence of truly ancient asexual species, this so-called 'Meselson effect' has been put forward as a means of confirming the complete loss of sexual reproduction. Very few attempts have, however, been made at quantifying the effect of sexual reproduction on the degree of divergence between gene copies in an asexual population. Here, I describe how asexual reproduction can be regarded as a special case of population subdivision. Using a slightly modified version of the standard two-deme structured coalescent, I derive the expected coalescence time for a pair of gene copies in an asexual population and show that the Meselson effect is compatible with low rates of sexual reproduction.

1. Introduction

In a randomly mating population of eukaryotic diploid organisms, the amount of neutral genetic divergence between two gene copies is expected to be the same irrespective of whether the two gene copies come from a single individual (i.e. the two alleles at a locus) or from different individuals, because of the random segregation of alleles. By contrast, eukaryotes that have abandoned sexual reproduction (i.e. meiosis) will no longer experience either segregation or recombination. This means that, if we look backwards in time, the two alleles at a locus in a diploid asexual organism cannot coalesce until sexual reproduction occurs (Nordborg & Krone, 2001). As a consequence, when there has been no sexual reproduction for very long periods of time, intra-individual allelic differences will vastly exceed those normally encountered in sexually reproducing organisms (Birky, 1996) and may even be larger than differences between species, a phenomenon known as the 'Meselson effect' (Judson & Normark, 1996).

* Corresponding author. Department of Cell and Organism Biology, Lund University, Sölvegatan 29, SE-22362, Lund, Sweden. Tel: +46 46 2227857. Fax: +46 46 147874. e-mail: alf.ceplitis@cob.lu.se

The existence of species that have persisted without sexual reproduction for very long periods of time – so-called 'ancient asexual scandals' – has been the subject of some controversy, because it is difficult to reconcile with current theories on the long-term evolutionary advantages of sex and recombination (Judson & Normark, 1996; Normark *et al.*, 2003). Many species suspected of being ancient asexuals are small and not easily studied in the field. The lack of sexual reproduction in such species is often inferred from circumstantial evidence (e.g. failure to observe males). Because sexual reproduction in putative ancient asexuals might be infrequent, essential aspects of their life-cycle might escape detection (Hurst *et al.*, 1992). To overcome these difficulties, it was suggested that the observation of extreme intra-individual allelic sequence divergence would confirm the lack of sexual reproduction in such species (Birky, 1996; Judson & Normark, 1996; Normark, 1999). Empirical evidence for the Meselson effect was first provided by Mark Welch & Meselson (2000), who showed that intra-individual differences between alleles at nuclear loci were extremely large in bdelloid rotifers. This finding was taken as convincing proof of the complete absence of sex in bdelloid rotifers for tens of millions of years. Similar studies have been performed on various

species believed to have lived without sex for millions of generations but have generally failed to show the same compelling results (Belshaw *et al*., 1999; Normark, 1999; Schön & Martens, 2003; but see Kuhn *et al*., 2001). It has been pointed out that failure to detect high levels of intra-individual allelic sequence divergence does not rule out complete asexuality because non-meiotic events, such as mitotic recombination or gene conversion, might have consequences similar to those of sexual reproduction (Birky, 1996; Butlin, 2002; Normark *et al*., 2003). However, it is not clear whether some (low) rate of sexual reproduction is compatible with the Meselson effect. In fact, little attention has been paid to the development of methods for more formal statistical analysis of genetic data obtained from asexual species.

The coalescent theory is a powerful tool for population genetic modelling (Nordborg, 2001). In particular, coalescent models incorporating population structure have been shown to be applicable to a wide range of genetic and evolutionary phenomena (e.g. geographic subdivision, balancing selection and diploidy) (Nordborg, 1997, 2001). The insight that asexual reproduction in a diploid organism leads to the isolation and divergence of the two homologous genomes within individuals implies that asexuality can be treated as a special kind of population subdivision. In this case, subdivision is not geographic but 'genomic'; each individual in a population harbours two distinct 'genomic subpopulations'. Sexual reproduction (i.e. segregation) will allow genes (or whole genomes if recombination is absent) to 'migrate' between the genomic subpopulations, reducing the divergence between them. In this sense, asexual reproduction thus most closely resembles the case of a balanced polymorphism, where recombination allows genes at a neutral locus to migrate between allelic classes maintained by balancing selection acting on a linked locus (e.g. Nordborg, 1997).

In this paper, I attempt to show how asexual reproduction can be modelled within the framework of the structured coalescent. Specifically, I derive the expected coalescence time for a pair of gene copies sampled from a population with low rates of sexual reproduction and show how the expected genetic diversity is affected. In addition, I briefly consider how the rate of sexual reproduction might be estimated in a population and discuss the validity of proposed tests of long-term asexuality based on intra-individual sequence divergence.

2. The model

(i) *Preliminaries*

The genealogy of a pair of gene copies in an asexual population can be modelled as a structured coalescent process with minor modifications. Consider first a population made up of N diploid individuals obeying a standard Wright–Fisher model of reproduction. In this population, there are thus $2N$ gene copies in total. The population can be viewed as divided into two haploid 'genomic subpopulations' of size $N_1 = N_2 = N$. Notice that the genomic subpopulations are not defined by the genealogy of a particular set of gene copies but are fixed abstract entities introduced for the purpose of modelling the genealogy. Hence, the two allelic gene copies from a single individual always belong, by definition, to two different genomic subpopulations, and every gene copy in the population belongs to either of the two genomic subpopulations, irrespective of the degree of sexuality in the population.

The aim is now to analyse the effect of segregational 'migration' between the genomic subpopulations. Consider, therefore, two selectively neutral gene copies (i.e. two pieces of non-recombining DNA) sampled from the population. The genealogy of the two gene copies may be described by a discrete-time Markov process with, as a first step, five possible states denoted (a, b), where a and b indicate the number of distinct gene copies in the first and second genomic subpopulations, respectively. It must be emphasized here that the two genomic subpopulations are, for all relevant purposes, completely equivalent. Thus, the terms 'first' and 'second' do not imply any particular order. The five possible states are thus $(1, 0)$, $(2, 0)$, $(1, 1)$, $(0, 2)$ and $(0, 1)$, where $(1, 0)$ and $(0, 1)$ are the ancestral states. By necessity, the subpopulation sizes are equal and the 'migration rate' is equal in both directions, so the model is completely symmetric. Consequently, the states $(1, 0)/(0, 1)$ and $(2, 0)/(0, 2)$ are equivalent and may be collectively denoted (1) and (2), respectively. The model can thus be reduced to one of three different states: (1), (2) and $(1, 1)$, where (1) denotes the ancestral state, (2) denotes two distinct genes in the same genomic subpopulations, and $(1, 1)$ denotes two distinct genes in different genomic subpopulations. A graphical representation of the genealogy is given in Fig. 1.

So far, the model is exactly like a standard two-deme structured coalescent as described by, for example, Nordborg (1997). However, as is, the model does not take into account the fact that, when two gene copies are found in different genomic subpopulations, the lineages leading back to their most recent common ancestor might at some point back in time coalesce into the same diploid individual (Fig. 1). In fact, two lineages in different genomic subpopulations (in different individuals) are expected to do this at a rate $1/N$. The transition probabilities from state $(1, 1)$ differ depending on whether the two gene copies are currently in two different individuals or in the same individual. This will affect the expected coalescence time for two gene copies sampled from

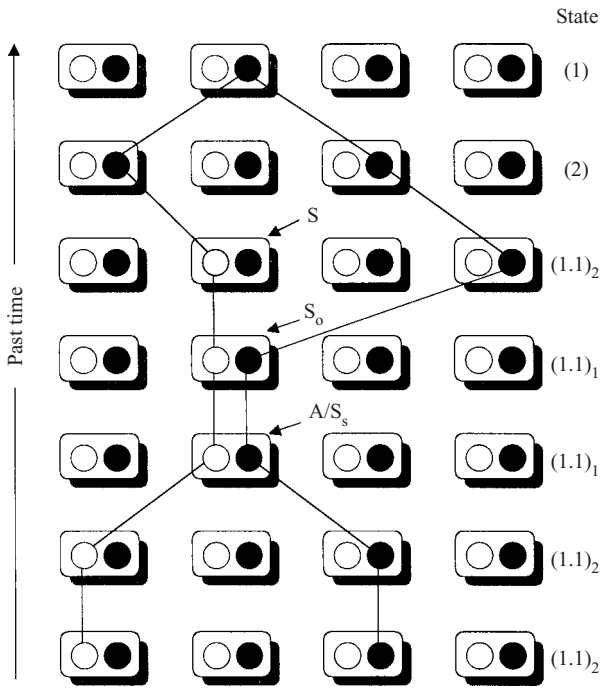

Fig. 1. Schematic representation of a population of a diploid asexual organism with rare sexual reproduction. Individuals are represented as rectangles. Both 'genomic subpopulations' (empty and filled circles) are represented in each diploid individual. The lines track the genealogy of a sample of two gene copies from the population. The individual labelled 'A/S_S' must be derived by either an asexual reproductive event or a sexual selfing event; the individual labelled 'S_O' must be derived from a sexual outcrossing event, whereas the individual labelled 'S' must be sexually derived, either by selfing or outcrossing. Other individuals might be derived either sexually or asexually. 'MRCA' indicates the most recent common ancestor of the two sampled genes. The leftmost column shows the different states in which the two gene copies occur in each generation. Many generations will pass between state transitions.

the population. To accommodate this in the model, the state (1, 1) is decomposed into two states: $(1,1)_2$, denoting two genes from different genomic subpopulations in two different individuals; and $(1,1)_1$, denoting two genes from different genomic subpopulations in the same individual. Thus, we now have a model with four different states: (1), (2), $(1,1)_2$, and $(1,1)_1$. Finally, let **R** be the matrix of transition probabilities r_{ij} $(i, j=1,\ldots, 4)$ between the different states.

(ii) *Coalescence time for a pair of genes under asexuality and outcrossing*

We consider first the situation when sexual reproduction occurs by outcrossing. To specify the transition probabilities of the matrix **R**, let ω be the probability that an individual is sexually produced and let $\lim_{N\to\infty} N\omega = \Omega$, so that $\omega = \Omega/N$. All parameter values are assumed to be constant over time. Ignoring terms of order N^{-2} and smaller yields the following matrix **R** of approximate transition probabilities is produced.

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{N} & 1-\frac{\Omega+1}{N} & \frac{\Omega}{N} & 0 \\ 0 & \frac{\Omega}{N} & 1-\frac{\Omega+1}{N} & \frac{1}{N} \\ 0 & \frac{\Omega}{2N} & \frac{\Omega}{2N} & 1-\frac{\Omega}{N} \end{pmatrix}. \quad (1)$$

The states are arranged in the order (1), (2), $(1,1)_2$ and $(1,1)_1$ from left to right.

If time is measured in units of $2N$ generations and $N\to\infty$ then the process converges to a continuous-time Markov process. Using standard analytical procedures (see Appendix), we obtain the expected coalescence times for a pair of gene copies

$$\begin{aligned} E[T(2)] &= 1+\frac{1}{2(2\Omega+1)}, \\ E[T(1,1)_2] &= 1+\frac{3\Omega+2}{2\Omega(2\Omega+1)}, \\ E[T(1,1)_1] &= 1+\frac{1}{\Omega}, \end{aligned} \quad (2)$$

where $E[T(i)]$ is the expected time to coalescence for two gene copies in state i. Notice that $E[T(1,1)_1]$ is the expected coalescence time for the two alleles at a locus sampled from a single individual. For two gene copies sampled from different individuals but otherwise at random, the expected coalescence time ($E[T]$) is

$$E[T] = \tfrac{1}{2}E[T(2)] + \tfrac{1}{2}E[T(1,1)_2] = 1+\frac{1}{2\Omega}, \quad (3)$$

because, under this sampling scheme, there is a 50% chance that the two gene copies are drawn from the same genomic subpopulation and a 50% chance that they are drawn from different genomic subpopulations.

Consider now what happens when sexual reproduction is common. The assumption $\lim_{N\to\infty} N\omega = \Omega$ that was initially made means that segregation events are rare (i.e. occur at a rate of the order N^{-1}). If we relax this assumption and allow sexual reproduction to be so frequent that, before a coalescence event happens, many segregation events have occurred then the process will behave like an unstructured coalescent process. Because of the symmetry of the model, the rate of the coalescent process when sexual reproduction is frequent is, in analogy to the standard two-deme structured coalescent, equal to 1 in units of $2N$ generations (cf. Nagylaki, 1980; Nordborg, 1997; Maruyama, 1972), which is the same as in a fully sexual, random mating population of N diploid individuals (Hudson, 1990; Nordborg, 2001).

https://doi.org/10.1017/S0016672303006487 Published online by Cambridge University Press

(iii) *Incorporating selfing*

The genealogical process with asexuality and rare sexual outcrossing is a special case of a more general model that allows a varying degree of selfing. Sexual reproduction by (partial) self-fertilization can be easily incorporated into the model. By letting f denote the probability that a sexually produced individual is derived from a selfing event, the matrix $\mathbf{R}_f$ gives the transition probabilities between the different states

$$\mathbf{R}_f=\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{N} & 1-\frac{\Omega+1}{N} & \frac{\Omega}{N} & 0 \\ 0 & \frac{\Omega}{N} & 1-\frac{\Omega+1}{N} & \frac{1}{N} \\ \frac{\Omega f}{2N} & \frac{\Omega(1-f)}{2N} & \frac{\Omega(1-f)}{2N} & 1-\frac{\Omega}{N}\left(1-\frac{f}{2}\right) \end{pmatrix}, \tag{4}$$

where the states are ordered as above. The expected coalescence times for two gene copies now become

$$\begin{aligned} E[T(2)]_f &= 1+\frac{1}{2(2\Omega+1)}-\frac{\Omega f}{2\Omega+1}, \\ E[T(1,1)_2]_f &= 1+\frac{3\Omega+2}{2\Omega(2\Omega+1)}-\frac{f(1+f)}{2\Omega+1}, \\ E[T(1,1)_1]_f &= 1+\frac{1}{\Omega}-f, \end{aligned} \tag{5}$$

and the expected coalescence time for two gene copies sampled at random from different individuals is

$$E[T]_f=1+\frac{1}{2\Omega}-\Phi, \text{ where } \Phi=\frac{f(\Omega+f+1)}{2(2\Omega+1)}. \tag{6}$$

As might be expected, selfing reduces the expected coalescence time for a pair of gene copies. The magnitude of the effect that selfing has depends, however, on the rate of sexual reproduction. In fully sexual populations, selfing decreases the effective population size and thus coalescence times, with a factor $1-f/2$ (Pollak, 1987; Schoen & Brown, 1991; Nordborg & Donnelly, 1997). In the present situation, notice that $\lim_{\Omega\to 0, f\to 1}\Phi=1$, which means that, when sex is rare, even obligate selfing will have practically no effect on the expected coalescence time for two randomly chosen gene copies. The same holds for two gene copies (alleles) taken from a single individual, as seen above.

(iv) *Genetic diversity in asexual populations*

If the total number of mutations along a branch in a genealogy is Poisson distributed with the mean $\theta=\mu t$, where μ is the mutation rate and t is the length of the branch (in some unit of time), the total number of mutational differences between two gene copies is 2θ, because they are separated by a divergence time of $2t$. In an asexual population with rare sexual outcrossing, the expected numbers of mutational differences between two randomly chosen gene copies (θ_a) and between two gene copies (alleles) from a single individual ($\theta_{a[I]}$) are

$$\theta_a=4\mu N\left(1+\frac{1}{2\Omega}\right), \tag{7a}$$

and

$$\theta_{a[I]}=4\mu N\left(1+\frac{1}{\Omega}\right), \tag{7b}$$

where time is now measured in units of one generation. In a fully sexual, randomly mating population of N diploid individuals, the expected number of mutational differences between two randomly chosen gene copies is $\theta_s=4N\mu$. This implies that genetic diversity in an asexual population is expected to be greater than in an equally sized sexual population, provided that sexual reproduction in the asexual population occurs at a rate in the order of N^{-1} or lower. Nevertheless, if $\Omega=1$, there is only a 1·5-fold difference between an asexual and a fully sexual population with respect to the number of mutational differences between two randomly chosen gene copies. Thus, regardless of the size of an asexual population, a single sexually produced recruit every generation is sufficient to give the asexual population a level of genetic diversity that, for all practical purposes, is likely to be indistinguishable from that of a fully sexual population. This conclusion agrees well with results from empirical studies on plants showing that average values of within-population genetic diversity measures do not differ significantly between sexual and partly asexual species (Hamrick & Godt, 1989; Table 3).

(v) *Estimating the rate of sexual reproduction*

Finally, we need an estimator of Ω, the scaled rate of sexual reproduction. Consider a situation in which both gene copies (alleles) are sequenced from several diploid individuals. Under an infinite-sites model of DNA sequence evolution, the expected number of differences per nucleotide site, $E[\pi]$, between two gene copies (sequences) is equal to θ (Li, 1997). Unlike the case of geographic subdivision, in which the migration rate can be estimated by comparing averaged estimates of θ for pairs of genes sampled from the same subpopulation and from the total population (e.g. Slatkin, 1991), there is, in this case, no way of knowing to which genomic subpopulation a given gene belongs. However, two genes sampled from the same individual always belong to different genomic subpopulations. Furthermore, for all pairwise comparisons of sequences sampled from different individuals,

https://doi.org/10.1017/S0016672303006487 Published online by Cambridge University Press

half will involve sequences from the same genomic subpopulation and half will involve sequences from different genomic subpopulations. Thus, we may, as an approximation, use the average number of observed nucleotide differences between pairs of sequences from within individuals, $\bar{\pi}_I$, to estimate $\theta_{a[I]}$, and the average number of observed nucleotide differences for all pairs of sequences from different individuals, $\bar{\pi}_D$, to estimate θ_a. From these observations, Ω may be estimated as

$$\hat{\Omega} = \frac{2\bar{\pi}_D - \bar{\pi}_I}{2(\bar{\pi}_I - \bar{\pi}_D)}. \quad (8)$$

It is assumed that there is equilibrium between drift and sexual reproduction, so that the estimated value of $\hat{\Omega}$ reflects the long-term, evolutionary stable value of the parameter Ω.

3. Discussion

As shown in the present study, the expected coalescence time for a pair of gene copies in an asexual population can be derived using a slightly modified two-deme structured coalescent model. This finding further extends the applicability of the structured coalescent, which has been used to model other evolutionary phenomena, such as geographic subdivision and balancing selection (Nordborg, 1997, 2001). The model described here considers only the effect of segregation on the genealogy of neutral genes. Sexual reproduction is often equated with recombination and, although the effect of segregation on the dynamics of deleterious mutations has been analysed (Antezana & Hudson, 1997), few studies have considered how neutral genetic diversity is affected by a limited rate of segregation. Inasmuch as segregation causes the random distribution of alleles at a locus among offspring, a restricted rate of segregation is the ultimate cause of enhanced allelic divergences in asexual organisms. Moreover, even though crossing-over and recombination are often required for proper separation of homologous chromosomes in meiosis (Davis & Smith, 2001), segregation can happen without concomitant recombination in a given genomic region (which is why linkage disequilibrium can be observed in sexually reproducing organisms). Recombination between strictly neutral loci will only cause different loci to segregate independently of each other. As a result, unlinked loci will have the same expected coalescence time, but the correlation between their respective genealogies will weaken with increasing recombination (e.g. Hudson, 1990). This is, of course, true for any degree of sexuality; the main consequence of asexual reproduction is a reduced effective rate of recombination.

In addition to segregation, there are several mechanisms – both meiotic and mitotic – that can bring about 'migration' between the genomic subpopulations and that might affect the genealogical process if they disrupt the symmetry and/or rate of the 'migration' model (cf. Nordborg, 1997). Asymmetry can be imposed by biased gene conversion or directional selection at linked sites (i.e. selection at a site at some recombination distance from the gene under consideration). Directional selection in combination with reduced recombination rates has, for example, been cited as a factor that should reduce within-population genetic diversity in asexual organisms (Charlesworth & Wright, 2001). However, if these processes either take place during meiosis (gene conversion) or depend on meiosis to take effect (selection at linked sites), they are also effectively limited by the rate of segregation. In fact, directional selection has no effect on intra-individual sequence divergence under compete asexuality (Birky, 1996). Thus, the consequences of meiotic gene conversion and directional selection will be of importance only when sexual reproduction is frequent. The same is true, as shown, for selfing: the probability that two gene copies in the same individual coalesces in the previous generation increases with selfing but the effect is far outweighed by a low rate of sexual reproduction (Eqn 6). By contrast, recombination and gene conversion during mitosis are independent of segregation and will have an effect similar to that of sexual reproduction (Birky, 1996). If such non-meiotic 'migration' events are common, the coalescence time for two gene copies might be significantly reduced even in the absence of sexual reproduction. Indeed, crossing-over during mitosis occurs at frequency of 10^{-3} to 10^{-2} relative to meiotic crossing-over (van Heemst & Heyting, 2000), so mitotic recombination might have important consequences when sex is rare. Both mitotic recombination and gene conversion have been invoked to explain the lack of high intra-individual sequence divergence in species believed to be ancient asexuals (Normark, 1999), even though the frequency with which such events take place in asexual species is not generally known.

Bengtsson recently studied the effect of asexual reproduction in a coalescent framework (Bengtsson, 2003). Although some aspects of his results are, in a general and qualitative sense, similar to those presented here (e.g. that asexuality has no effect on the coalescence time of a pair of gene copies unless sexual reproduction is rare), the model that he uses does not give a true representation of the genealogical process in an asexual population. In particular, Bengtsson regards a pair of gene copies in an asexual population as occurring in one of two states (Bengtsson, 2003) – in the same individual or in different individuals – and considers the probability that two gene copies drawn from different individuals will coalesce in the previous generation to be $1/2N$. This is clearly incorrect because

the probability of coalescence, and thus the expected coalescence time, for two genes taken from different individuals depends on whether the two genes come from – using the terminology of the present paper – the same or different genomic subpopulations (see the Model section). That this is the case is obvious from both the expected gene phylogenies in asexual organisms (Birky, 1996, p. 429; Judson & Normark, 1996) and empirical data (Mark Welch & Meselson, 2000). Consequently, the model of Bengtsson (2003) cannot account for observed patterns of genetic diversity in asexual organisms.

In the first attempt to quantify the Meselson effect, Birky (1996) showed that the expected number of substitutions between alleles within diploid individuals of an asexual species, k, is $2\mu(2N+t)$, where $2N$ is the expected coalescence time for the two alleles in the ancestral sexual population and t is the time, in generations, since the origin of asexuality and during which there has been no sexual reproduction (Birky, 1996, p. 428). Giving a numerical example, Birky (1996) calculated that, with typical values of $4N\mu = 0{\cdot}01$ and $\mu = 5 \times 10^{-9}$ per site, complete absence of sexual reproduction for 10 million years (i.e. $t = 10^7$ assuming one generation per year) would result in an expected number of substitutions per site between alleles (k) of 0·11, compared with 0·01 in a fully sexual species. Birky (1996) conjectured that occasional sexual reproduction should drastically reduce the coalescence time of two gene copies from a single individual and proposed that $k \geqslant 0{\cdot}10$ constitutes good evidence for long-term asexual reproduction. However, with the same values of N and μ as in the example above, an intra-individual sequence divergence of $k = 0{\cdot}11$ is expected if $\Omega = 0{\cdot}1$ (i.e. if there is one sexually derived individual every tenth generation, which is to say that k is equivalent to $\theta_{a[I]}$ in Eqn 7b). Thus, ancient asexuality is not the only plausible explanation of high levels of intra-individual sequence divergence; the Meselson effect may manifest itself even with non-negligible rates of sexual reproduction.

In a landmark study of the Meselson effect, Mark Welch & Meselson (2000) sequenced both copies of four genes from single individuals of four species of bdelloid rotifers. In most cases, highly divergent and presumably allelic gene copies were found within single individuals, whereas gene copies from different individuals (and species) were either equally divergent or much more similar. For example, phylogenetic analysis of sequences from the gene *hsp82* revealed that each sequence belonged to either of two ancient lineages, and that representatives of both lineages were found in the individuals of the two species *Macrotrachela quadricornifera* and *Philodina roseola* (Mark Welch & Meselson, 2000). The sequence data from the *hsp82* gene lend itself to a preliminary estimation of Ω, as outlined in the previous section. Again, $\bar{\pi}_I$ is the average observed difference between gene copies from a single individual, and $\bar{\pi}_D$ corresponds to the average observed difference between gene copies from different individuals (the fact that different individuals belong to different species is of no relevance to the procedure). In this case, $\bar{\pi}_I = 49$ and $\bar{\pi}_D = 23{\cdot}6$ (values are from uncorrected percentage difference at fourfold degenerate sites; see Table 2 in Mark Welch & Meselson, 2000). Using Eqn 8, the estimated scaled rate of sexual reproduction in these bdelloid rotifer species is $\hat{\Omega} = 0{\cdot}145$. This result suggests that the rotifer sequence data are compatible with the production of roughly one sexual recruit every seventh generation, a rate of sexual reproduction that, if true, is clearly incompatible with the status of the bdelloid rotifers as ancient asexuals. Indeed, recent analysis of sequence data has indicated the occurrence of historical recombination events in several putative ancient asexuals (Gandolfi *et al*., 2003), casting some doubt on the absence of sexual reproduction in these species. More work, both empirical and theoretical, is clearly needed in this field. In particular, the statistical properties of the estimator of Ω need to be carefully studied before it can be reliably applied to data of the kind provided by Mark Welch & Meselson (2000). Better estimates of Ω might be found by using, for example, Bayesian inference methods based on summary statistics (Beaumont *et al*., 2002). Still, as indicated by the results of the present study, DNA sequence data might not provide the sought-after conclusive evidence of ancient asexuality.

Appendix

The derivation of the expected coalescence times for a pair of gene copies follows standard procedures (e.g. Nordborg, 1997). Given the matrix **R** of transition probabilities, with the different states arranged as described in the Model section (Eqn 1), the amount of time that two gene copies spend in state i ($i > 1$) before entering another state is a geometrically distributed random variable T_i. The probability that T_i is larger than some amount of time τ, is

$$P(T_i > \tau) = (r_{ii})^{\tau}, \tag{A1}$$

where r_{ii} is a diagonal element of **R**. If time is measured in units of $2N$ generations and $N \to \infty$, we get

$$\begin{aligned}
\lim_{N\to\infty} P\left(T_2 > 1 - \frac{1}{2N}(2\Omega+1)\right)^{2N\tau} &= e^{-(2\Omega+1)\tau},\\
\lim_{N\to\infty} P\left(T_3 > 1 - \frac{1}{2N}(2\Omega+1)\right)^{2N\tau} &= e^{-(2\Omega+1)\tau},\\
\lim_{N\to\infty} P\left(T_4 > 1 - \frac{1}{2N}2\Omega\right)^{2N\tau} &= e^{-2\Omega\tau},
\end{aligned} \tag{A2}$$

https://doi.org/10.1017/S0016672303006487 Published online by Cambridge University Press

for $i=2,\ldots,4$ [i.e. states (2), $(1,1)_2$ and $(1,1)_1$]. This means that T_i has an exponential distribution in the limit. The expected amount of time, $E[T_i]$, that is spent in state i is thus

$$\begin{aligned} E[T_2] &= \frac{1}{2(\Omega+1)}, \\ E[T_3] &= \frac{1}{2(\Omega+1)}, \\ E[T_4] &= \frac{1}{2\Omega}, \end{aligned} \tag{A3}$$

for $i=2,\ldots,4$. Once the two gene copies leave state i, the (conditional) probability that they enter another state j is $r_{ij} \div (1-r_{ii})$. The coalescence time for two gene copies currently in state i is given by the expected time spent in state i plus the conditional probability of entering state j times the expected time spent in j and summed over all j different from i; in the present case, this is

$$E[T(i)] = E[T_i] + \sum_{j=2, j\neq i}^{4} \left(\frac{r_{ij}}{1-r_{ii}} E[T_j] \right), \tag{A4}$$

where $E[T(i)]$ is the expected coalescence time for two gene copies in state i. The expected coalescence times for two gene copies in states $i=2,\ldots,4$ can thus be found from the following set of equations:

$$\begin{aligned} E[T(2)] &= \frac{1}{2(\Omega+1)} + \frac{\Omega}{\Omega+1} E[T(1,1)_2], \\ E[T(1,1)_2] &= \frac{1}{2(\Omega+1)} + \frac{\Omega}{\Omega+1} E[T(2)] \\ &\quad + \frac{1}{\Omega+1} E[T(1,1)_1], \\ E[T(1,1)_1] &= \frac{1}{2\Omega} + \tfrac{1}{2} E[T(2)] + \tfrac{1}{2} E[T(1,1)_2]. \end{aligned} \tag{A5}$$

Solving this system of equations yields the expected coalescence times given in Eqn 2. Substituting matrix $\mathbf{R}_f$ for $\mathbf{R}$ gives the expected coalescence times with selfing (Eqn 5).

I am grateful to M. Nordborg, H. Innan and two anonymous reviewers for helpful suggestions and comments on the manuscript. I also thank B. Bengtsson and M. Lascoux for interesting discussions on the subject.

References

Antezana, M. A. & Hudson, R. R. (1997). Before crossing over: the advantages of eukaryotic sex in genomes lacking chiasmatic recombination. *Genetical Research* **70**, 7–25.

Beaumont, M. A., Zhang, W. & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.

Belshaw, R., Quicke, D. L. J., Völkl, W. & Godfray, H. C. J. (1999). Molecular markers indicate rare sex in a predominantly asexual parasitoid wasp. *Evolution* **53**, 1189–1199.

Bengtsson, B. O. (2003). Genetic variation in organisms with sexual and asexual reproduction. *Journal of Evolutionary Biology* **16**, 189–199.

Birky, C. W., Jr (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **144**, 427–437.

Butlin, R. K. (2002). The costs and benefits of sex: new insights from old asexual lineages. *Nature Reviews Genetics* **3**, 311–317.

Charlesworth, D. & Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics and Development* **11**, 685–690.

Davis, L. & Smith, G. R. (2001). Meiotic recombination and chromosome segregation in *Schizosaccharomyces pombe*. *Proceedings of the National Academy of Sciences of the USA* **98**, 8395–8402.

Gandolfi, A., Sanders, I. R., Rossi, V. & Menozzi, P. (2003). Evidence of recombination in putative ancient asexuals. *Molecular Biology and Evolution* **20**, 754–761.

Hamrick, J. L. & Godt, M. J. (1989). Allozyme diversity in plant species. In *Plant Population Genetics, Breeding, and Genetic Resources* (ed. A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir), pp. 45–53. Sunderland, MA: Sinauer.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology* (ed. D. Futuyma & J. Antonovics), pp. 1–43. Oxford: Oxford University Press.

Hurst, L. D., Hamilton, W. D. & Ladle, R. J. (1992). Covert sex. *Trends in Ecology and Evolution* **7**, 144–145.

Judson, O. P. & Normark, B. B. (1996). Ancient asexual scandals. *Trends in Ecology and Evolution* **11**, 41–46.

Kuhn, G., Hijri, M. & Sanders, I. R. (2001). Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* **414**, 745–748.

Li, W.-H. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer.

Mark Welch, D. & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211–1215.

Maruyama, T. (1972). Some invariant properties of a geographically structured finite population: distribution of heterozygotes under irreversible mutation. *Genetical Research* **20**, 141–149.

Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *Journal of Mathematical Biology* **9**, 101–114.

Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics* **146**, 1501–1514.

Nordborg, M. (2001). Coalescent theory. In *Handbook of Statistical Genetics* (ed. D. J. Balding, M. Bishop & C. Cannings), pp. 179–212. Chichester: John Wiley & Sons.

Nordborg, M. & Donnelly, P. (1997). The coalescent process with selfing. *Genetics* **146**, 1185–1195.

Nordborg, M. & Krone, S. M. (2001). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Theoretical Population Genetics* (ed. M. Slatkin & M. Veuille), pp. 194–232. Oxford: Oxford University Press.

Normark, B. B. (1999). Evolution in a putatively ancient asexual lineage: recombination and rapid karyotype change. *Evolution* **53**, 1458–1469.

Normark, B. B., Judson, O. P. & Moran, N. A. (2003). Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society* **79**, 69–84.

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

Schoen, D. J. & Brown, A. H. D. (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the USA* **88**, 4494–4497.

Schön, I. & Martens, K. (2003). No slave to sex. *Proceedings of the Royal Society of London Series B* **270**, 827–833.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genetical Research* **58**, 167–175.

van Heemst, D. & Heyting, C. (2000). Sister chromatid cohesion and recombination in meiosis. *Chromosoma* **109**, 10–26.