Extending Coalescent Theory to Autotetraploids

Extending Coalescent Theory to Autotetraploids

B. Arnold,[1] K. Bomblies, and J. Wakeley
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT We develop coalescent models for autotetraploid species with tetrasomic inheritance. We show that the ancestral genetic process in a large population without recombination may be approximated using Kingman's standard coalescent, with a coalescent effective population size $4N$. Numerical results suggest that this approximation is accurate for population sizes on the order of hundreds of individuals. Therefore, existing coalescent simulation programs can be adapted to study population history in autotetraploids simply by interpreting the timescale in units of $4N$ generations. We also consider the possibility of double reduction, a phenomenon unique to polysomic inheritance, and show that its effects on gene genealogies are similar to partial self-fertilization.

POLYPLOIDY, which results from whole-genome duplication, is a significant evolutionary force throughout the tree of life. It is particularly widespread in higher plants but also occurs in fishes, amphibians, reptiles, insects, and even a mammal (Sexton 1980; Gallardo *et al.* 1999; Leggatt and Iwama 2003; Gregory and Mable 2005). In plants, estimates of the proportion of angiosperm species that have experienced genome doubling during their evolutionary history vary from 30 to 100% (Stebbins 1950; Grant 1981; Masterson 1994; Ciu *et al.* 2006), and polyploidy is thought to be a potent mechanism of sympatric speciation (Wood *et al.* 2009).

Polyploids can arise via interspecific hybridization (allopolyploids) or intraspecific genome duplication (autopolyploids), *e.g.*, through the fusion of unreduced gametes (Stebbins 1947). Allopolyploids, born from the union of distinct genomes, frequently exhibit bivalent pairing and disomic inheritance. As a result, the duplicated chromosome sets are only partially homologous (homeologous) and follow separate evolutionary paths if they do not pair and recombine. Conversely, autotetraploids contain four nondiverged sets of chromosomes that are fully homologous. During meiosis, these homologs may form multivalents or bivalents with random chromosome pairing, resulting in tetrasomic inheritance in both cases.

doi: 10.1534/genetics.112.140582
Manuscript received March 26, 2012; accepted for publication June 1, 2012
Available freely online through the author-supported open access option.
Supporting information is available online at www.genetics.org/content/suppl/2012/06/18/genetics.112.140582.DC1.
[1]Corresponding author: Department of Organismic and Evolutionary Biology, Harvard University, 22 Divinity Ave., Cambridge, MA 02138. E-mail: barnold@oeb.harvard.edu

Autotetraploids were once thought to be vanishingly rare compared to allopolyploids, but they are much more common than previously suspected (Soltis *et al.* 2007). Numerous wild plant species have been demonstrated to exhibit tetrasomic inheritance with random chromosome pairing, despite forming only bivalents at meiosis I (Supporting Information, Table S1 and references therein; Soltis *et al.* 2007). However, allele segregation in autopolyploids can become much more complex than that in diploids when chromosomes form multivalents from crossing over with more than one homolog during meiosis, as this may lead to double reduction. Double reduction occurs when multivalents are resolved such that segments of sister chromatids migrate to the same pole at meiosis I, allowing, for example, an ABBB genotype to produce AA gametes (Haldane 1930; Mather 1935). Among cytologically characterized natural autotetraploid plants, multivalent formation is less common than bivalent pairing but is still present in enough species to merit attention in theoretical models.

The body of literature on theoretical population genetics of polyploids has grown but is still very small in comparison to the work done on diploids (Bever and Felber 1992; Otto and Whitton 2000). Studies have characterized equilibrium genotype proportions (Haldane 1930), genetic drift, and levels of genetic variation (Wright 1938; Moody *et al.* 1993), as well as population structure of autotetraploid populations (Ronfort *et al.* 1998; Luo *et al.* 2006). A few studies have considered the effects of double reduction on the equilibrium frequencies of neutral and deleterious alleles (Crow 1954; Butruille and Boiteux 2000). The gene genealogical or coalescent approach to population genetics (Hudson 1983; Tajima 1983) has proved a useful framework for interpreting

genetic variation. Coalescent models have been applied to data from tetraploid species, but the justification for this has not been elucidated. The ability to extract all the information from a set of DNA sequences collected from natural populations will help answer questions about autopolyploid evolution. For example, do most autopolyploids experience severe bottlenecks from the formation event? How many independent formation events do most autopolyploid species experience? Is there gene flow between ploidy levels? How old are these autopolyploid populations and how evolutionarily stable is tetrasomic inheritance? What does genome structure look like in natural populations, in terms of evolutionarily important parameters such as the population mutation and recombination rates?

Despite the increasing awareness of autopolyploidy, the burgeoning advances in DNA sequencing technology, and the utility of these data for studying evolution, few studies have analyzed nuclear DNA sequence variation in natural populations of autotetraploids (Tiffin and Gaut 2001; Jørgensen *et al.* 2011; St. Onge *et al.* 2012). Here we develop a coalescent model for autotetraploids to facilitate the analysis of DNA sequence data. Specifically, we derive the coalescent effective population size (Sjödin *et al.* 2005) for autotetraploid species. We consider both double reduction and the possibility of partial selfing. Our mathematical results hold in the limit as the population size tends to infinity, but we show that they are numerically very accurate when the population size is only moderately large (in the hundreds). These results provide a mathematical framework to explicitly model DNA sequence evolution in autotetraploid populations, which may be employed to estimate mutation and recombination rates, infer ancestral demography, and detect various types of selection from DNA sequence data sets of diploids. In short, we show how standard coalescent models or simulations may be applied to autotetraploids with only minor modification, thus allowing for detailed predictions to be made about patterns of genetic variation and for population history and the evolutionary forces acting on natural populations to be inferred from DNA sequence data.

Theory

Kingman (1982a,b) gave a formal proof of the existence of what he called the "n-coalescent"—now simply coalescent or Kingman's coalescent—as the ancestral genetic process for a sample from a large haploid population. Specifically, for a genetic locus at which all variation is selectively neutral and there is no intralocus recombination, the genetic ancestry of a sample of size n may be modeled as a process in which each pair of lineages ancestral to the sample "coalesces" with rate equal to one. Derivations of the coalescent begin with the computation of single-generation probabilities (*i.e.*, of coalescence) and then proceed by taking a limit as the population size (N) tends to infinity, rescaling time in units proportional to N generations so that a coalescence rate of one per pair of lineages is obtained. Coalescent models have been extended to include population subdivision and migration, changes in population size over time, recombination, and natural selection. Efficient software is available to simulate samples of genetic data (Hudson 2002; Ewing and Hermisson 2010).

Kingman's derivation of the coalescent is valid only for haploid population models (this includes the diploid monoecious Wright–Fisher model, because it can be reduced to a haploid model). In particular, Kingman assumed that genetic lineages are "exchangeable" as in the haploid population models of Cannings (1974). The general formalism for treating diploids or other (nonexchangeable) population structures was developed by Möhle (1998a,b,c). It has been applied in a variety of situations to show that Kingman's coalescent is robust to deviations from Kingman's original assumptions, but also to describe an augmented set of ancestral genetic processes that are closely related to Kingman's coalescent; see Wakeley (2008) for a review. Here we use Möhle's technique to derive the ancestral genetic process at a single neutral locus without recombination in an autotetraploid species.

Sample size of two chromosomes

Consider a Wright–Fisher population of N autotetraploid, monoecious individuals that create gametes via tetrasomic inheritance with no recombination (*i.e.*, no double reduction, which is considered later). That is, an individual with four distinct alleles at a locus produces $\binom{4}{2} = 6$ kinds of gametes, each with equal frequency. Generations are nonoverlapping. Each of the N offspring that form the next generation is created by the union of two gametes sampled randomly with replacement from all possible gametes of the parental generation. Thus each individual is produced by selfing with probability $1/N$. Due to the added structure in which each gamete contains two distinct parental copies of each locus, and in contrast to the diploid monoecious case, samples of chromosomes from a tetraploid species are not exchangeable. In particular, the conditional probability of coalescence depends on whether two genetic lineages are within the same individual or in separate individuals. For example, without recombination and double reduction, two lineages cannot coalesce in the immediately previous generation if they are within the same individual and came from the same gamete.

We can construct a single-generation transition matrix **P** for an ancestral process that accounts for the specific details of tetrasomic inheritance by setting up an absorbing Markov chain that has three states for a sample size of $n = 2$: (1) two distinct lineages within the same individual, (2) two lineages in separate individuals, and (3) a single lineage that is the common ancestor of the two lineages. As we trace them back in time, the two lineages may travel between states one and two until they ultimately coalesce. Transition probabilities are calculated by conditioning on two pieces of information: whether lineages came from the same gamete in the previous generation and the particular pattern of ancestry.

For instance, if both lineages are in the same individual (state 1), they coalesce in the previous generation if they came from different gametes (probability 2/3) produced by the same parent (probability $1/N$) and trace back to the same chromosome in that parent (probability 1/4) (see Figure 1). Thus, $\mathbf{P}_{1,3} = 1/6N$. This is in agreement with Wright (1938), who showed that the "proportion of unlike pairs of genes" decays by a factor of 5/6 for self-fertilizing autotetraploids. If two lineages are in separate individuals (state 2), the probability of coalescence must be computed for all possible patterns of shared ancestry that potentially result in identity by descent (Table S2), with the marginal probability of coalescence being $\mathbf{P}_{2,3} = 1/4N$ after simplification.

Applying the same logic to calculate the probabilities of other possible transitions, we get the following 3×3 single-generation transition matrix $\mathbf{P}$ with states 1, 2, and 3 represented by rows one, two, and three, respectively:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{3} + \frac{1}{2N} & \frac{2}{3}\left(1 - \frac{1}{N}\right) & \frac{1}{6N} \\ \frac{3}{4N} & \left(1 - \frac{1}{N}\right) & \frac{1}{4N} \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix describes the exact, discrete-time ancestral process for the two lineages. As in Kingman's coalescent, we seek a continuous-time approximation that will be accurate when the population size is large. Specifically, we take the limit $N \to \infty$, with time rescaled by N, to test whether the ancestral limit process converges to Kingman's coalescent.

Möhle (1998a) obtained a useful convergence result for Markov processes with two timescales such as the one described by the matrix above. Since $\mathbf{P}$ contains terms that become increasingly different in the limit $N \to \infty$, we can use Möhle's result to construct a continuous-time approximation. We rewrite $\mathbf{P}$ in three parts, such that

$$\mathbf{P} = \mathbf{F} + \frac{\mathbf{S}}{N} + o\left(\frac{1}{N}\right), \tag{1}$$

where $\mathbf{F} = \lim_{N\to\infty}\mathbf{P}$ and $\mathbf{S} = \lim_{N\to\infty} N(\mathbf{P} - \mathbf{F})$. Matrix $\mathbf{F}$ contains the "fast" events that occur frequently on the timescale of the original discrete-time process (*i.e.*, the movement of lineages from within to between gametes), whereas matrix $\mathbf{S}$ contains the "slow" events (*i.e.*, coalescence). The terms in $\mathbf{S}/N$ become very small as N tends to infinity. These events occur on the timescale of N generations. While Möhle's result allows for additional terms, of $o(1/N)$, which tend to zero more quickly than $1/N$, in our case these terms are equal to zero.

If the matrix $\mathbf{E} = \lim_{t\to\infty}\mathbf{F}^t$ exists, the continuous-time approximation to our ancestral process involves the fast events instantaneously reaching their equilibrium ($\mathbf{E}$), after which they enter the slower process of coalescence described by rate matrix $\mathbf{G} = \mathbf{ESE}$ (Möhle 1998a). More formally, Möhle's Theorem 1 (Möhle 1998a) states that the rescaled ancestral process converges to $\lim_{N\to\infty}\mathbf{P}^{[Nt]} = \mathbf{E}e^{t\mathbf{G}}$. Together matrices $\mathbf{E}$ and $\mathbf{G}$ describe the rescaled ancestral process, with time measured in units of N generations. Applying this theorem to our discrete-time matrix above, we have

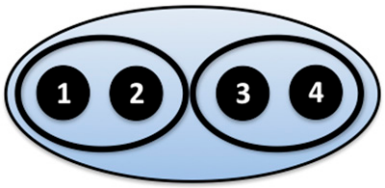


Figure 1 A diagram of a tetraploid individual, showing the two diploid gametes that united to create it. If chromosomes are sampled without replacement, after the first lineage is sampled, with probability 2/3 we sample the second lineage from a different gamete, which came from the same parent with probability $1/N$ (where N is the population size), assuming random mating. Conditional on these events, these two sampled lineages coalesce with probability 1/4.

$$\mathbf{F} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} \frac{1}{2} & -\frac{2}{3} & \frac{1}{6} \\ \frac{3}{4} & -1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

such that

$$\mathbf{E} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 0 & -\frac{1}{4} & \frac{1}{4} \\ 0 & -\frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}.$$

Here, we see that if lineages2 start out in state 1 (within the same individual), they may remain in that state if they came from the same gamete ($\mathbf{F}_{1,1}$) or different gametes from the same parent ($\mathbf{S}_{1,1}$). However, our continuous-time approximation for large populations shows that the number of generations the lineages remain in state 1 is negligible on the timescale of N generations, so they immediately travel to separate individuals (from state 1 to state 2 since $\mathbf{E}_{1,2} = 1$). Once in state 2, the pair of lineages enter the coalescence process given by $\mathbf{G}$, which is a simple exponential process in which coalescence occurs with rate equal to 1/4 (on the timescale of N generations), which agrees with forward-time models of autotetraploid populations (Moody *et al.* 1993). If time is rescaled again, by the constant factor 4, so that 1 unit of time is equal to $4N$ generations, then the rate of coalescence is one, just as in Kingman's coalescent. That is, we

Table 1 Markov states that account for all possible configurations of four lineages in an autotetraploid population, with two types of absorbing states to assess the relative probabilities of single- and multiple-coalescence events

State	Description
1	All four lineages within the same individual
2	Three lineages within the same individual and one lineage in a separate individual
3	Two pairs of lineages, each within the same individual
4	Two lineages within the same individual and the other two each in separate individuals
5	All four lineages in separate individuals
6	Three lineages remain after a single-coalescence event
7	One or two lineages remain after a multiple-coalescence event

have shown for a sample size of 2 that the coalescent process for an autotetraploid species without double reduction converges to the same limiting ancestral process as that of a haploid population model when the coalescent effective size is defined as $N_e = 4N$.

Extension to larger samples

Convergence to Kingman's coalescent for a sample of size 2 does not guarantee convergence for larger samples. It must also be true that pairwise coalescent events dominate the limiting ancestral process (Möhle and Sagitov 2001). In the case of an autotetraploid without double reduction, we must check that multiple coalescent events between lineages within a single individual are negligible in the limit $N \to \infty$. For example, four lineages within a single individual will coalesce in two pairs in the immediately previous generation with probability $1/6N$, which is of the same order of magnitude as the rate of pairwise coalescence. In fact, such events become negligible in the limit because four lineages in a single individual will be overwhelmingly more likely to be descended from two pairs of lineages in two different individuals. We show this by briefly repeating our previous analysis, using Möhle's (1998a) technique, or Equation 1, but for a sample of $n = 4$.

We construct this more complicated ancestral process by defining a new absorbing Markov chain with seven states that include all possible configurations of four lineages. Transient states 1–5 account for all possible ways lineages can be distributed within and between individuals (Table 1). States 6 and 7 are both absorbing, with the former representing single, pairwise coalescence events and the latter defined to include all possible multiple-coalescence events. States 6 and 7 are absorbing in the sense that here we are concerned only with the process during which there are four ancestral lineages. However, the coalescent process resumes on the remaining ancestral lineages, if more than one remain, with a new transition matrix appropriate for the smaller sample size. As in the derivation for $n = 2$, transition probabilities between states can be calculated by conditioning on patterns of ancestry. All possible patterns for four lineages, along with their associated conditional transition probabilities, are shown in Table S3.

We obtain a new **P** that is now the 7×7 single-generation transition matrix for a sample of size $n = 4$ (not shown). The continuous-time approximation for a large population is obtained by applying Möhle's (1998a) theorem, decomposing **P** into three separate matrices that contain events that occur on different timescales, as done above. In the limit as $N \to \infty$, with time rescaled in N generations, we obtain matrices **E** and **G** that describe the ancestral process:

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \\ 0 & 0 & 0 & 0 & -3/2 & 3/2 & 0 \end{bmatrix}$$

Similarly to the process for $n = 2$, there is an instantaneous adjustment of the sample to a state in which all lineages are in separate individuals (here state 5), independent of the starting state. The sample then enters the continuous-time process given by **G**, in which the rate of single coalescence events, on the timescale of N generations, is six times greater than in the case of $n = 2$, or $6 \times (1/4) = 3/2$. This is identical to Kingman's coalescent, in which coalescence times are exponentially distributed and occur with rate equal to the number of pairs of lineages: $\binom{4}{2}$.

Our result is analogous to the one obtained by Möhle (1998b) for a diploid dioecious population. In the diploid case, even though two lineages in a single individual cannot coalesce in the previous generation (thus violating the exchangeability assumption of Kingman's coalescent), lineages quickly assume a state in which each is in a separate individual, and the long-term coalescence rate is equal to one per pair of lineages when time is measured in units of $2N$ generations. In the tetraploid monoecious case, although

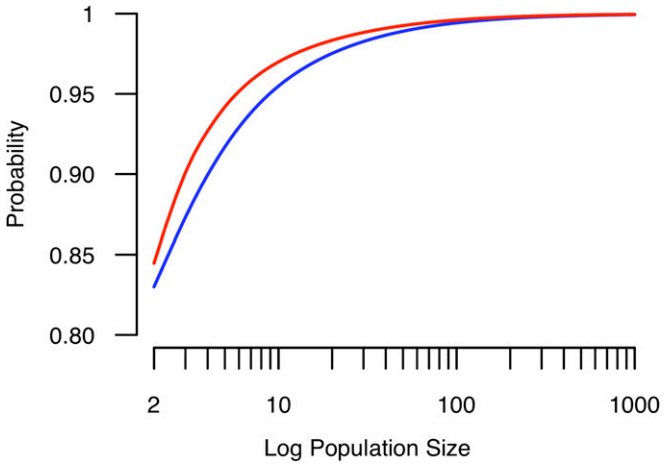


Figure 2 The probability of a single coalescence event, conditional on an absorbing event, given the process started in state 1 (blue) or state 5 (red) for N up to 1000 autotetraploid individuals.

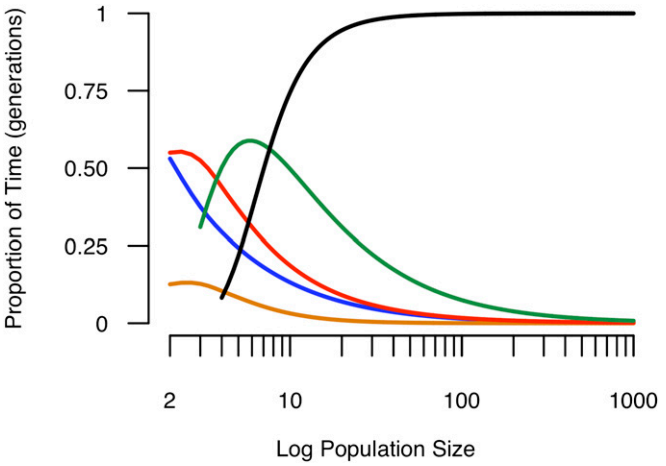


Figure 3 The expected proportion of generations the Markov chain spends in state 1 (blue), state 2 (orange), state 3 (red), state 4 (green), or state 5 (black) for a range of population sizes. Here the process starts in state 1 but spends a vast majority of its time in state 5, even for small populations. These proportions were calculated from matrix **N** (see text). Note that state 4 is possible only when $N \geq 3$ and state 5 is possible only when $N \geq 4$.

lineages may travel together in gametes for some number of generations without coalescing, they are overwhelmingly more likely to become scattered such that each is in a separate individual, and the long-term coalescence rate is equal to one per pair of lineages when time is measured in units of $4N$ generations.

Accuracy for finite N

Although the continuous-time approximation applies in the limit as population size tends to infinity, we want to know the validity of this approximation for smaller, finite populations. All the information about the ancestral process is contained in the single-generation transition martix **P**, so we can analyze the dynamics of this discrete-time process for a range of population sizes and compare them to the continuous-time approximation that allows only single coalescent events. We investigate the accuracy of two key features of the limiting ancestral process: that coalescence events occur predominantly between pairs of lineages and that the majority of the ancestral process before coalescence is spent in state 5 (*i.e.*, with all lineages in separate individuals, making them exchangeable).

We can do this using standard theory of absorbing Markov chains, for example in Chap. 11 of Grinstead and Snell (1997), by writing the transition matrix in canonical block form,

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

in which $\mathbf{Q}$ is a 5×5 matrix of transition probabilities among transient (noncoalescent) states, $\mathbf{R}$ is a 5×2 matrix of transition probabilities from transient to asborbing (coalescent) states, $\mathbf{0}$ is a 2×5 zero matrix, and $\mathbf{I}$ is the identity matrix (in this case 2×2). Then, for each starting state, the 5×5 matrix inverse $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ contains the expected numbers of generations spent in each transient state before absorption and the 5×2 matrix product $\mathbf{NR}$ contains the probabilities of absorption in each absorbing state.

In terms of how genetic data are typically sampled, there are two extreme starting states for the process: state 1 or state 5, in which a biologist samples all chromosomes from a single autotetraploid individual or one chromosome from each of four individuals. Given these starting states, the probability of a single coalescence event (absorbing to state 6 rather than state 7) is plotted in Figure 2 for a range of population sizes, indicating rapid convergence to an ancestral process involving predominantly pairwise coalescence events. Likewise, Figure 3 shows that the fraction of time spent in state 5 (when all lineages are in separate individuals) approaches one quickly as the population size increases. In Figure 3 the sample is assumed to start in state 1, which may be considered the farthest from the transient equilibrium (state 5) of the limiting ancestral process. From Figures 2 and 3, we conclude that the limiting ancestral process for a sample of size $n = 4$ should be quite accurate as long as the population size is $> \sim 100$.

Extension to arbitrary sample size n

An exact calculation for larger sample sizes is not practical given the complexity of the model with $n = 4$. However, a strong heuristic derivation can be made based on the fact that when $n \ll N$, events that occur with probability $O(1)$ or $O(1/N)$ per generation will dominate the ancestral process. In short, if n lineages are in fewer than n individuals, the most probable events are $O(1)$ and these send the lineages into different individuals. If instead each lineage is in a separate individual, the most probable events are $O(1/N)$ and these bring lineages into the same individual, possibly to

coalesce. Due to the $O(1)$ transition probabilities, the chance that n lineages will ever be in fewer than $n-1$ individuals (other than when they are originally sampled) is negligible if $n \ll N$.

Thus, starting with n lineages in n individuals, we can accurately summarize the process using a Markov chain with just three states: (1) n lineages are within $n-1$ individuals, (2) n lineages remain in n individuals, and (3) $n-1$ lineages are arrived at by a coalescent event (see Table S4). We have

$$\mathbf{P}=\begin{bmatrix} \frac{1}{3}+O\left(\frac{1}{N}\right) & \frac{2}{3}+O\left(\frac{1}{N}\right) & O\left(\frac{1}{N}\right) \\ \left(\binom{2n}{2}-n\right)\frac{3}{16N}+O\left(\frac{1}{N^2}\right) & 1+\left(n-\binom{2n}{2}\right)\frac{1}{4N}+O\left(\frac{1}{N^2}\right) & \left(\binom{2n}{2}-n\right)\frac{1}{16N}+O\left(\frac{1}{N^2}\right) \\ 0 & 0 & 1 \end{bmatrix}.$$

Applying Möhle's (1998a) theorem, we obtain matrices $\mathbf{E}$ and $\mathbf{G}$,

$$\mathbf{E}=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{G}=\begin{bmatrix} 0 & -\binom{n}{2}\frac{1}{4} & \binom{n}{2}\frac{1}{4} \\ 0 & -\binom{n}{2}\frac{1}{4} & \binom{n}{2}\frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix},$$

which demonstrates that the ancestral process, starting either in state 1 or in state 2, corresponds to Kingman's standard n-coalescent if time is measured in units of $4N$ generations.

An ancestral process with double reduction

Autotetraploids contain four sets of homologous chromosomes, and crossovers may potentially occur among any of them. When more than one crossover arises per chromosome, they may involve different pairing partners and create multivalents at metaphase I. Depending on how chromosomes segregate, double reduction may occur at a particular locus if there is a crossover between this locus and the centromere (Mather 1935, 1936; Crow 1954). Retrospectively, these lineages are automatically identical by descent (*i.e.*, they coalesce) in the previous generation. This is qualitatively similar to the case of partial selfing studied by Nordborg and Donnelly (1997) and Möhle (1998a), in which lineages within a single individual may coalesce in the immediately previous generation if the individual was produced by selfing.

Multivalent formation is a necessary precursor for double reduction (Mather 1935). Most established autotetraploids form bivalents at meiosis (see Table S1) almost certainly because multivalents are associated with aneuploid gametes and thus reduced fitness of progeny (reviewed in Comai 2005). Double reduction is thus a phenomenon primarily of newly formed autotetraploids that have not adapted to a genome-doubled state, but individuals capable of correctly segregating multivalents could theoretically be selected for (Comai 2005). Depending on how meiotic mechanisms evolve in a particular autotetraploid species, double reduction may not occur over long enough periods of evolutionary time (*i.e.*, on the coalescent timescale of N generations) to have a significant effect. Nonetheless, the presence of double reduction in at least some natural autopolyploids means that it merits consideration. We use Möhle's technique here as well, with $n=2$, to study the effects of double reduction on the coalescent process.

Consider a locus at some distance from the centromere, such that recombination may occur between them. For simplicity, we assume that recombination does not occur within the locus under consideration. Following Stift *et al.* (2008), the frequency of double reduction (α) has a theoretical maximum value of 1/6, assuming that chromosomes always form quadrivalents and one crossover occurs between the locus and the centromere. With probability 1/3, the recombined chromosomes migrate to the same pole during meiosis I (assuming all chromosome pairs are equally likely to segregate), and the probability that the two sister chromatids also migrate to the same pole during meiosis II is 1/2. Double reduction can be less than this maximum if recombination between the locus and the centromere does not occur in every multivalent association (*i.e.*, the locus is distal to the centromere) or if chromosomes form bivalents with some probability at meiosis I. Thus, α may range from 0 to 1/6 and will differ between loci.

To study an ancestral process for $n=2$ that includes double reduction, we define an absorbing Markov chain with the same three states used previously. The resulting matrix looks very similar to the one for tetrasomic inheritance:

$$\mathbf{P}=\begin{bmatrix} \frac{1-\alpha}{3}+\frac{1}{2N} & \frac{2}{3}-\frac{2}{3N} & \frac{\alpha}{3}+\frac{1}{6N} \\ \frac{3}{4N} & 1-\frac{1}{N} & \frac{1}{4N} \\ 0 & 0 & 1 \end{bmatrix}.$$

This shows that double reduction increases the probability of coalescence from state 1 by an amount $\alpha/3$. Since double reduction does not affect lineages that came from separate gametes, the transition probabilities for state 2 are unchanged. When in state 1, if lineages came from the same gamete in the previous generation, which occurs with probability 1/3, there is a chance α that they were sister chromatids in the immediately previous generation and thus coalesce. Importantly, the chance of this novel transition when lineages are in state 1 does not depend on the population size N. There is no reason to suppose that α is small, and we assume that it is a constant when we take the limit $N \to \infty$. Using Möhle's (1998a) theorem to obtain a continuous-time approximation on the timescale of N generations, we have

$$\mathbf{E} = \begin{bmatrix} 0 & \frac{2}{2+\alpha} & \frac{\alpha}{2+\alpha} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and}$$

$$\mathbf{G} = \begin{bmatrix} 0 & \frac{3}{(\alpha+2)^2} - \frac{2}{\alpha+2} & \frac{3\alpha}{2(\alpha+2)^2} + \frac{1}{2(\alpha+2)} \\ 0 & \frac{3}{2(\alpha+2)} - 1 & \frac{1}{4} + \frac{3\alpha}{4(\alpha+2)} \\ 0 & 0 & 0 \end{bmatrix}.$$

Unlike the previous models above, coalescence is now possible in both the fast and slow processes. On the limiting timescale of N generations, lineages may instantaneously coalesce instead of moving directly to state 2. This will create an association of alleles within individuals. Once in state 2, lineages will remain in this state for a majority of their ancestry but coalesce at a rate faster than 1/4 when time is measured in units of N generations. Thus double reduction decreases the long-term coalescent N_e. In addition, a single exponentially distributed rate of coalescence (*i.e.*, the Kingman coalescent) is not sufficient to capture all the dynamics of an ancestral process with double reduction; with time rescaled in units proportional to N generations, lineages may coalesce instantaneously if they were sampled from the same individual.

As previously mentioned, these effects are also observed in the coalescent process with partial self-fertilization (Nordborg and Donnelly 1997; Möhle 1998b). However, the two models are not identical for tetraploids. In File S1, we extend our model of coalescence for autotetraploids without double reduction to included partial selfing, with probability s. We find that the maximum rate of double reduction ($\alpha = 1/6$) produces the same ancestral process as one with $s = 1/4$, but that there are slight quantitative differences between the two models. Figure 4 shows the relationship between s and α, in terms of the parameter values that give the same coalescent N_e. The relationship is only slightly nonlinear; though mechanistically distinct, selfing and double reduction have very similar effects on patterns and levels of genetic variation.

The coalescent with double reduction for abitrary n

Consider a sample of n alleles from a population such that k individuals contain two alleles, l individuals contain three alleles, m individuals contain four alleles, and $n - 2k - 3l - 4m$ alleles are in separate individuals. Unlike the extension to arbitrary n for the model above, coalescence events may occur with $O(1)$ if double reduction is possible. Thus, the instantaneous adjustment of the sample involves both the movement of lineages from within to between individuals and coalescence events. After this instantaneous adjustment, the remaining lineages are in separate individuals and coalesce with probability $O(1/N)$.

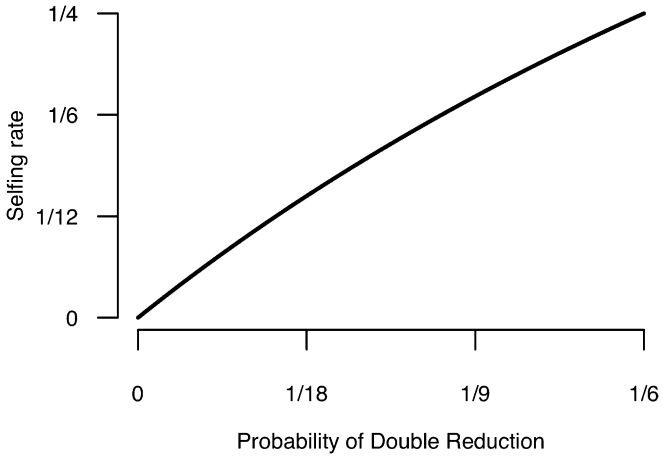


Figure 4 The slightly nonlinear relationship between double reduction and self-fertilization. The maximum probability of double reduction (1/6) corresponds to a selfing rate of 1/4.

The number of instantaneous coalescence events depends on the sample configuration, *i.e.*, the number of lineages within an individual. Following Nordborg and Donnelly (1997), for each of the k individuals that contain two lineages, the number of instantaneous coalescence events is $X \sim \text{Binomial}(k, \alpha/(\alpha + 2))$, with the probability of coalescence calculated above from the analysis with $n = 2$. For the l individuals that contain three lineages, the number of instantaneous coalescence events is $Y \sim \text{Binomial}(l, 3\alpha/(\alpha + 2))$: two of the lineages must come from the same gamete and coalesce with probability α, but if they do not coalesce with probability $1 - \alpha$, then they coalesce in previous generations from double reduction with probability $(\alpha/(\alpha + 2))$. Thus, the overall probability of coalescence is $\alpha + (1 - \alpha)(\alpha/(\alpha + 2)) = 3\alpha/(\alpha + 2)$. Applying the same logic to the m individuals with four lineages (see File S1), the number of coalescence events is $\mathbf{Z} = (Z_1, Z_2, Z_3) \sim \text{Multinomial}(m, \mathbf{p} = (p_1, p_2, p_3))$. Here, Z_1 is the number of single coalescence events that occur with probability $p_1 = 12\alpha(1-\alpha)/(\alpha+2)^2$, Z_2 is the number of double coalescence events that occur with probability $p_2 = 9\alpha^2/(\alpha+2)^2$, and Z_3 is the number of times no lineages coalesce (*i.e.*, four lineages within an individual get separated into four distinct individuals) with probability $p_3 = 4(\alpha-1)^2/(\alpha+2)^2$. The numbers of lineages that remain after this instantaneous adjustment, $n - X - Y - Z_1 - 2Z_2$, are in separate individuals and enter the slow process of coalescence with rate $\binom{n-X-Y-Z_1-Z_2}{2}(1/4 + 3\alpha/4(\alpha+2))$ if time is measured in units of N generations. The coalescence rate calculated here is the same as above for $n = 2$ but applied to each pair of remaining lineages.

Discussion

Autotetraploids with tetrasomic inheritance have long been considered vanishingly rare but now are increasingly recognized

as a common phenomenon in plant evolution (Soltis *et al.* 2007). Many established polyploid species or populations have been shown to have tetrasomic inheritance (Table S1). Nuclear DNA sequence data will provide invaluable insight into many unknown aspects of autopolyploid evolution, such as the process of formation and establishment from relatively few individuals that are at least partially reproductively isolated from their diploid progenitor. Here, we extend the coalescent, a widely used model in population genetics and phylogenetics, to autotetraploid populations to aid in the analysis of DNA sequence data sets from these species.

Our results show that, although tetrasomic inheritance creates additional configurations of lineages that have unique probabilities of coalescence compared to those for diploid organisms, the ancestral process for autotetraploids without double reduction converges to Kingman's haploid coalescent model with time rescaled by $4N$ generations. Intuitively, this convergence occurs because in a large, panmictic population ancestral lineages quickly get separated to different individuals such that the sample spends a majority of its history in a configuration in which lineages are exchangeable.

Simulating data with the tetrasomic inheritance model without double reduction would produce genealogies like those generated from diploid and haploid models, such as Hudson's ms (Hudson 2002), with the exception that the timescale of the process must be interpreted differently. The time to the most recent common ancestor (T_{MRCA}) of a pair of lineages in an autotetraploid population is exponentially distributed with a mean of one when time is measured in units of $4N$ generations. Thus a given value of T_{MRCA} is interpreted as $4N \times T_{\mathrm{MRCA}}$ rather than $2N \times T_{\mathrm{MRCA}}$ as in diploid coalescent models. From this it follows that, as has been previously demonstrated (Moody *et al.* 1993), autotetraploids are expected to have twice the levels of genetic variation as diploids for a given demographic size. However, many demographic or biological factors may lead to departures from this expectation, such as population history, the distribution of offspring per individual, or nonrandom mating.

A useful application of the coalescent is to infer ancestral demography. Little is known about the characteristics of the bottlenecks associated with the formation of an autotetraploid lineage, which can arise through the union of unreduced gametes produced by diploids. Since autotetraploid formation events are relatively rare, but potentially on the order of the genic mutation rate (Ramsey and Schemske 1998), entire populations may be founded by a small number of individuals, resulting in a severe genetic bottleneck. However, the severity of this bottleneck likely varies among cases as certain environmental factors and alleles may greatly affect the rate of unreduced gamete formation in diploids (Ramsey and Schemske 1998); higher rates can lead directly to repeated autotetraploid formation or promote gene flow from diploids to tetraploids. Multiple formation events or gene flow from diploid gene pools may increase the effective population size during autotetraploid formation and reduce the severity of the bottleneck. Several studies have documented multiple origins in autotetraploids (Soltis *et al.* 1989; Wolf *et al.* 1990; Ptacek *et al.* 1994). Alternatively, gene flow among ploidy levels may occur via interploidy hybrids (*i.e.*, triploids) if they have nonzero fertility and produce some euploid gametes (Felber and Bever 1997; Husband 2004).

The utility of the models developed here depends on the specific demographic history of the autotetraploid population. If the present-day autotetraploid race traces back to relatively few individuals (*i.e.*, $\sim$10), then the sample size n, or number of ancestral lineages, is not much smaller than the population size N. Since $n \ll N$ is a critical assumption of the Kingman coalescent, the continuous-time models developed here in the context of large population size may not be applicable. Simulated gene genealogies produced by standard coalescent programs such as Hudson's ms may not look like the actual pattern of ancestry that contains multifurcations, a likely gene tree structure of small populations that affects predicted patterns of genetic variation. Special simulation programs may be developed that can accommodate extreme population size crashes, for example as in the metapopulation model of Wakeley and Aliacar (2001). Alternatively, the distribution of pairwise coalescence and linkage disequilibrium can be used to quantify the severity of the bottleneck, such as Li and Durbin's (2011) pairwise coalescent model, if multiple mergers are a likely feature of the underlying genealogy. Simulations should be done to see whether these models are robust to such large population size reductions.

If mating is not random (*i.e.*, self-fertilization occurs) or if the autotetraploids form multivalents and exhibit double reduction, the ancestral process does not converge to Kingman's simple model. In these cases, Kingman's coalescent does not fully capture the relatively complicated ancestry of the sample of lineages because the coalescent process cannot be described by a single exponential distribution; coalescent events may occur instantly with a nonzero probability even as population size tends to infinity. For sample sizes greater than two, simultaneous multiple mergers become an issue. For instance, if four lineages are within the same individual that was created by a selfing event, a simultaneous multiple merger occurs in the previous generation with probability $s/6$, where s is the selfing rate (see File S1 for details). This probability is independent of population size and thus does not tend to zero as $N \to \infty$. For a coalescent model with double reduction and four lineages within the same individual, as $N \to \infty$ gametes likely come from different parents, unlike in the selfing model. However, a simultaneous multiple merger occurs at the locus of interest if both gametes that formed the individual were produced by double-reduction events with probability α^2. Thus, these ancestral processes are more complex than the Kingman coalescent since coalescent events may happen on two separate timescales and simultaneous multiple mergers likely occur.

The coalescent model with double reduction is similar to the many-demes model, with its "scattering phase" and "collecting phase" (Wakeley 1999), and to the coalescent with partial selfing (Nordborg and Donnelly 1997; Mohle 1998a). It is interesting to note, however, that a maximum of two $O(1)$ coalescence events may occur in the instantaneous adjustment for four lineages sampled from within a single tetraploid individual. This results from the fact that with probability of $O(1)$ the two gametes that form each individual came from different parents, and only each pair that came from the same gamete may coalesce via double reduction. For finite N, more coalescence events are possible if gametes originated from the same parent, but this has probability $O(1/N)$ and thus does not happen in the instantaneous adjustment. Tetrasomic inheritance thus creates a genetic structure that has a mathematically distinct ancestral process.

However, there are two observations that suggest that, at least in plants, self-fertilization or double reduction may rarely affect results: first, of the established autotetraploid plant species that have been documented and studied, multivalent formation (and thus the possibility of double reduction) and self-fertilization are both rare. For example, among 24 species for which tetrasomic inheritance has been molecularly confirmed, chromosome associations have also been examined in 11. Among these, only 2 closely related species commonly form multivalents. Two other species form multivalents only rarely, while the remainder show exclusively bivalent associations at meiosis (Table S1). Thus in most of these examples, double reduction would be rare or absent. Self-fertilization is also rare. Although selfing can in theory promote the establishment of tetraploids by helping avoid minority cytotype exclusion (Rodriguez 1996), it is very rare among polyploid species in nature (Stebbins 1947). The majority of plant species we identified in the literature that have tetrasomic inheritance are obligately outcrossing (see Table S1 notes for references). Selfing rates are therefore generally zero for most autotetraploids. Thus, the simple tetrasomic coalescent model may be widely applicable.

In conclusion, our results demonstrate that Kingman's coalescent is robust to tetrasomic inheritance, making existing coalescent models applicable for analyzing population genomic data collected from natural autotetraploid populations that exhibit this mode of inheritance. These models will greatly facilitate the study of the evolutionary forces acting on these organisms. However, standard coalescent simulators cannot be used to interpret these data if the autotetraploids self-fertilize at an appreciable rate or if some loci experience double reduction, the latter being verified by cytology and segregation studies.

Literature Cited

Bever, J., and F. Felber, 1992 The theoretical population genetics of autopolyploids. Oxf. Surv. Evol. Biol. 8: 185–217.

Butruille, D. V., and L. S. Boiteux, 2000 Selection–mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. Proc. Natl. Acad. Sci. USA 97: 6608–6613.

Cannings, C., 1974 The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv. Appl. Probab. 6: 260–290.

Ciu, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis *et al.*, 2006 Widespread genome duplications throughout the history of flowering plants. Genome Res. 16: 738–749.

Comai, L., 2005 The advantages and disadvantages of being polyploidy. Nat. Rev. Genet. 6: 836–846.

Crow, J. F., 1954 Random mating with linkage in polysomics. Am. Nat. 88: 431–434.

Ewing, G., and J. Hermisson, 2010 MSMS: a coalescent simulation program including recombination, demographic structure, and selection at a single locus. Bioinformatics 26: 2064–2065.

Felber, F., and J. Bever, 1997 Effect of triploid fitness on the coexistence of diploids and tetraploids. Biol. J. Linn. Soc. Lond. 60: 95–106.

Gallardo, M., J. Bickham, R. Honeycutt, R. Ojeda, and N. Köhler, 1999 Discovery of tetraploidy in a mammal. Nature 401: 341.

Grant, V., 1981 *Plant Speciation*. Columbia University Press, New York.

Gregory, T. R., and B. K. Mable, 2005 Polyploidy in animals, pp. 427–517 in *The Evolution of the Genome*, edited by T. R. Gregory. Elsevier, San Diego.

Grinstead, C. M., and J. L. Snell, 1997 *Introduction to Probability*, Ed. 2. American Mathematical Society, Providence, RI.

Haldane, J., 1930 Theoretical genetics of autopolyploids. J. Genet. 22: 359–372.

Hudson, R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. 23: 183–201.

Hudson, R., 2002 Generating samples under a Wright–Fisher neutral model. Bioinformatics 18: 337–338.

Husband, B., 2004 The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. Biol. J. Linn. Soc. Lond. 82: 537–546.

Jørgensen, M., D. Ehrich, R. Schmickl, M. Koch, and A. Brysting, 2011 Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). BMC Evol. Biol. 11: 346.

Kingman, J. F. C., 1982a The coalescent. Stoch. Proc. Appl. 13: 235–248.

Kingman, J. F. C., 1982b On the genealogy of large populations. J. Appl. Probab. 19A: 27–43.

Leggatt, R. A., and G. K. Iwama, 2003 Occurrence of polyploidy in the fishes. Rev. Fish Biol. Fish. 13: 237–246.

Li, H., and R. Durbin, 2011 Inference of human population history form individual whole-genome sequences. Nature 475: 493–496.

Luo, Z. W., Z. Zhang, R. Zhang, M. Pandey, O. Gailing *et al.*, 2006 Modeling population genetic data in autotetraploid species. Genetics 172: 639–646.

Masterson, J., 1994 Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264: 421–424.

Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. J. Genet. 30: 53–78.

Mather, K., 1936 Segregation and linkage in autotetraploids. J. Genet. 32: 287–314.

Möhle, M., 1998a A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. Adv. Appl. Probab. 30: 493–512.

Möhle, M., 1998b Coalescent results for two-sex population models. Adv. Appl. Probab. 30: 513–520.

Möhle, M., 1998c Robustness results for the coalescent. J. Appl. Probab. 35: 438–447.

Möhle, M., and S. Sagitov, 2001 A classification of coalescent processes for haploid exchangeable population models. Ann. Appl. Probab. 29: 1547–1562.

Moody, M., L. Mueller, and D. Soltis, 1993 Genetic variation and random drift in autotetraploid populations. Genetics 134: 649–657.
Nordborg, M., and P. Donnelly, 1997 The coalescent process with selfing. Genetics 146: 1185–1195.
Otto, S. P., and J. Whitton, 2000 2000 Polyploid incidence and evolution. Annu. Rev. Genet. 34: 401–437.
Ptacek, M., H. Gerhardt, and R. Sage, 1994 Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, Hyla versicolor. Evolution 48(3): 898–908.
Ramsey, J., and D. Schemske, 1998 Pathways, mechanisms, and rates of polyploidy formation in flowering plants. Annu. Rev. Ecol. Syst. 29: 467–501.
Rodriguez, D. J., 1996 A model for the establishment of polyploidy in plants. Am. Nat. 147(1): 33–46.
Ronfort, J., E. Jenczewski, T. Bataillon, and F. Rousset, 1998 Analysis of population structure in autotetraploid species. Genetics 150: 921–930.
Sexton, O., 1980 Polyploidy in animal evolution: summary, pp. 379–381 in *Polyploidy, Biological Relevance*, edited by W. H. Lewis. Plenum, New York.
Sjödin, P., I. Kaj, S. Krone, M. Lascoux, and M. Nordborg, 2005 On the meaning and existence of an effective population size. Genetics 169: 1061–1070.
Soltis, D. E., P. S. Soltis, and B. D. Ness, 1989 Chloroplast-DNA variation and multiple origins of autopolyploidy in Heuchera micrantha (Saxifragaceae). Evolution 43(3): 650–656.
Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson *et al.*, 2007 Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? Taxon 56: 13–30.
Stebbins, G., 1947 Types of polyploids: their classification and significance. Adv. Genet. 1: 403–429.
Stebbins, G., 1950 *Variation and Evolution in Plants*. Columbia University Press, New York.
Stift, M., C. Berenos, P. Kuperus, and P. H. van Tienderen, 2008 Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. Genetics 179: 2113–2123.
St. Onge, K., J. Foxe, J. Li, H. Li, K. Holm *et al.*, 2012 Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Sheperd's Purse (Capsella bursa-pastoris). Mol. Biol. Evol. 29(7): 1721–1733.
Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437–460.
Tiffin, P., and B. Gaut, 2001 Sequence diversity in the tetraploid Zea perennis and the closely related diploid Z. diploperennis: insights from four nuclear loci. Genetics 158: 401–412.
Wakeley, J., 1999 Non-equilibrium migration in human history. Genetics 153: 1863–1871.
Wakeley, J., 2008 *Coalescent Theory: An Introduction*. Roberts, Greenwood Village, CO.
Wakeley, J., and N. Aliacar, 2001 Gene genealogies in a metapopulation. Genetics 159: 893–905.
Wolf, P., D. Soltis, and P. Soltis, 1990 Chloroplast-DNA and allozymic variation in diploid and autotetraploid Heuchera grossulariifolia (Saxifragaceae). Am. J. Bot. 77(2): 232–244.
Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon *et al.*, 2009 The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA 106: 13875–13879.
Wright, S., 1938 The distribution of gene frequencies in population of polyploids. Proc. Natl. Acad. Sci. USA 24: 372–377.

Communicating editor: Y. S. Song

GENETICS

Supporting Information
www.genetics.org/content/suppl/2012/06/18/genetics.112.140582.DC1

Extending Coalescent Theory to Autotetraploids

B. Arnold, K. Bomblies, and J. Wakeley

DOI: 10.1534/genetics.112.140582

Table S1 Plant species with confirmed tetrasomic inheritance

Species	Chromosome associations (metaphase I)	Reference(s)
Acacia nilotica		1
Actinidia chinensis		2
Allium nevii		3
Arabidopsis lyrata		4
Biscutella laevigata		5, 6
Centaurea jacea	Bivalents	7
Chrysanthemum boreale	Bivalents	8
Dioscorea trifida		9
Epilobium angustifolium	Mostly Bivalents, Some quadrivalents	10, 11
Heuchera grossulariifolia	Mostly Bivalents, Some quadrivalents	12
Heuchera micrantha		13
Lotus corniculatus	Bivalents	14, 15
Maclura pomifera		16
Medicago sativa		17
*Paspalum notatum***	Mostly quadrivalents	18
Paspalum simplex	Mostly quadrivalents	19
Prunus spinosa	Bivalents	20
Rorippa amphibia		21
Rorippa sylvestris		21
Rutidosis leptorrhynchoides		22
Thymus praecox		23
Tolmeia menziesii	Bivalents	24
Turnera ulmifolia	Bivalents	25, 26
Vaccinium corymbosum	Bivalents	27, 28

This table lists species that have been demonstrated to be genetically tetrasomic, meaning that four alleles segregate at single loci. It is meant to show that numerous species have been demonstrated to have tetrasomic inheritance, even if cytologically diploidized, and is not meant to be an exhaustive list.

** *Paspalum notatum* has tetrasomic inheritance at most markers, but in apmoctic lines has disomic inheritance around the apospory (apomixis) locus.

References**:**

1. Mandal, A.K., Ennos, R.A. and Fagg, C.W. (1994) Mating system analysis in a natural population of *Acacia nilotica* subspecies *leiocarpa*. *Theor Appl Genet* 89:931-935.
2. Huang, H., Dane, F., Wang, Z., Jiang, Z., Huang, R. and Wang, S. (1997) Isozyme inheritance and variation in *Actinidia*. *Heredity* 78: 328-336.
3. Rieseberg, L.H. and Doyle, M.F. (1989). Tetrasomic segregation in the naturally occurring autotetraploid *Allium nevii* (*Alliaceae*). *Hereditas* 111: 31-36.
4. Mable, B.K., Beland, J., and Di Berardo, C. (2004) Inheritance and dominance of self-incompatibility alleles in polyploid *Arabidopsis lyrata. Heredity* 93: 476-486.
5. Manton I. (1934) The problem of *Biscutella laevigata* L. *Z. Indukt. Abstammungs- Vererbungsl.* 67: 41–57.
6. Manton I. (1937) The problem of *Biscutella laevigata* L. II. The evidence from meisois. *Ann. Bot*. 1: 439–462.
7. Hardy, O.J., Vanderhoeven, S., De Loose, M. and Meerts, P. (2000) Ecological, morphological and allozymic differentiation between diploid and tetraploid knapweeds (*Centaurea jacea*) from a contact zone in the Belgian Ardennes. *New Phytol*. 146: 281-29
8. Watanabe, K. (1983) Studies on the control of diploid-like meiosis in polyploidy taxa of *Chrysanthemum. Theor Appl Genet* 66:9-14.
9. Bousalem, M., Arnau, G., Hochu, I., Arnolin, R., Viader, V., Santoni, S. and David, J. (2006) Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the *Dioscoreae*. *Theor Appl Genet* 113: 439-451.
10. Mosquin, T. (1967) Evidence for Autopolyploidy in *Epilobium angustifolium* (Onagraceae). *Evolution* 21: 713-719.
11. Husband, B.C. and Schemske, D.W. (1997) The Effect of Inbreeding in Diploid and Tetraploid Populations of *Epilobium angustifolium* (Onagraceae): Implications for the Genetic Basis of Inbreeding Depression. *Evolution* 51: 737-746.

12. Wolf, P.G., Soltis, P.S., and Soltis, D.E. (1989). Tetrasomic inheritance and chromosome pairing behaviour in the naturally occurring autotetraploid *Heuchera grossulariifolia* (Saxifragaceae). *Genome* 32 : 655 -659.
13. Soltis, D.E. and Soltis, P.S. (1989) Tetrasomic inheritance in *Heuchera micrantha* (Saxifragaceae). *J Hered* 80: 123-126.
14. Dawson, C.D.R. (1941) Tetrasomic inheritance in *Lotus corniculatus* L. *J Genet* 42: 49-73.
15. Fjellstrom, R.G., Beuselinck, P.R. and Steiner, J.J. (2001) RFLP marker analysis supports tetrasomic inheritance in *Lotus corniculatus* L. *Theor Appl Genet* 102:718-725
16. Laushman, R.H., Schnabel, A., and Hamrick, J.L. (1990). Electrophoretic Evidence for tetrasomic inheritance in the dioecious tree *Maclura pomifera* (Raf.) Schneid. *J Hered*, 87: 469-473.
17. Quiros, C.F. (1982) Tetrasomic segregation for multiple alleles in alfalfa. *Genetics* 101; 117-127.
18. Stein, J., Quarin, C.L., Matínez, E.J., Pessino, S.C., and Ortiz, J.P.A. (2004) Tetraploid races of *Paspalum notatum* show polysomic inheritance and preferential chromosome pairing around the apospory-controlling locus. *Theor Appl Genet* 109:186-191.
19. Pupilli, F., Caceres, M.E., Quarin, C.L. and Arcioni, S. (1997) Segregation analysis of RFLP markers reveals a tetrasomic inheritance in apomictic *Paspalum simplex*. *Genome* 40: 822-828.
20. Leinemann, L. (2000) Inheritance analysis of isozyme phenotypes in tetraploid species using single plant progenies. An example in black thorn (*Prunus spinosa* L.). *Forest Genetics* 7: 205-209.
21. Stift, M., Berenos, C., Kuperus, P. and van Tienderen, P.H. (2008) Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: A general procedure applied to *Rorippa* (Yellow Cress) microsatellite data. *Genetics* 179: 2113-2123.
22. Brown, A.H.D. and Young, A.G. (2000) Genetic diversity in tetraploid populations of the endangered daisy *Rutidosis leptorrhynchoides* and implications for its conservation. *Heredity* 85: 122-129.
23. Landergott, U., Naciri, Y., Schneller, J.J., and Holderegger, R. (2006) Allelic configurations and polysomic inheritance of highly variable microsatellites in tetraploid gynodioecious *Thymus praecox* agg. *Theor Appl Genet* 113: 453-465.
24. Soltis, D.E. and Rieseberg, L.H. Autopolyploidy in *Tolmiea menziesii* (Saxifragaceae): Genetic insights from enzyme electrophoresis. *Am J Bot* 73: 310-318.
25. Tamari, F., Khosravi, D., Hilliker, A.J. and Shore, J.S. (2005) Inheritance of spontaneous mutant homostyles in *Turnera subulata x krapovickasii* and in autotetraploid *T. scabra* (Turneraceae). *Heredity* 94: 207-216.
26. Fernández, A. (1987). Estudios cromosómicos en *Turnera* y *Piriqueta* (*Turneraceae*). *Bonplandia 6*: 1-21.
27. Jelenkovic, G., and Hough, L.F. (1970). Chromosome associations in the first meiotic division in three tetraploid clones of *Vaccinium corymbosum* L. *Can* J *Genet Cytol* 12: 316–324.
28. Krebs, S.L. and Hancock, J.F. (1989). Tetrasomic inheritance of isoenzyme markers in the highbush blueberry *Vaccinium corymbosum* L. *Heredity* 63: 11–18.

Table S2 Patterns of ancestry used to calculate transition probabilities for the ancestral process of 2 lineages sampled from an autotetraploid population. Balls represent autotetraploid individuals, and lines are the transmission of gametes. From left to right, the columns containing probabilities represent the probability of observing the pattern of ancestry and the conditional transition probabilities for the Markov chain.

Pattern	P{Pattern}	p_{11}	p_{12}	p_{13}
	$\frac{1}{N}$	$\frac{5}{6}$		$\frac{1}{6}$
	$\frac{N-1}{N}$	$\frac{1}{3}$	$\frac{2}{3}$	

Pattern	P{Pattern}	p_{21}	p_{22}	p_{23}
	$\frac{1}{N^3}$	$\frac{3}{4}$		$\frac{1}{4}$
	$\frac{N-1}{N^3}$		1	
	$4 \bullet \frac{N-1}{N^3}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
	$2 \bullet \frac{N-1}{N^3}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
	$2 \bullet \frac{(N-1)(N-2)}{N^3}$		1	
	$4 \bullet \frac{(N-1)(N-2)}{N^3}$	$\frac{3}{16}$	$\frac{3}{4}$	$\frac{1}{16}$
	$\frac{(N-1)(N-2)(N-3)}{N^3}$		1	

Note: For example, for the pattern of ancestry in which two autotetraploids are half-sibs, the probability of coalescence is 1/16 because, conditional on this pattern, there is a 1/4 chance that the two lineages in separate individuals came from the same parent and a 1/4 chance that they originated from the same chromosome. For the case in which two autotetraploids are full-sibs, the probability of coalescence is twice as great since it may occur in each of the two parents.

Table S3 Patterns of ancestry used to calculate transition probabilities for the ancestral process of 4 lineages sampled from an autotetraploid population. Balls represent autotetraploid individuals, and lines are the transmission of gametes. From left to right, the columns containing probabilities represent the probability of observing the pattern of ancestry and the conditional transition probabilities for the Markov chain.

Pattern	P{Pattern}	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}
	$\frac{1}{N}$	$\frac{1}{6}$					$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{N-1}{N}$			1				

Pattern	P{Pattern}	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}	p_{27}
	$\frac{1}{N^3}$	$\frac{1}{8}$					$\frac{5}{8}$	$\frac{1}{4}$
	$\frac{N-1}{N^3}$		$\frac{1}{2}$				$\frac{1}{2}$	
	$4 \bullet \frac{N-1}{N^3}$	$\frac{1}{32}$	$\frac{1}{8}$	$\frac{3}{8}$			$\frac{13}{32}$	$\frac{1}{16}$
	$2 \bullet \frac{N-1}{N^3}$		$\frac{1}{4}$	$\frac{3}{8}$			$\frac{3}{8}$	
	$2 \bullet \frac{(N-1)(N-2)}{N^3}$		$\frac{1}{4}$		$\frac{1}{2}$		$\frac{1}{4}$	
	$4 \bullet \frac{(N-1)(N-2)}{N^3}$		$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{2}$		$\frac{3}{16}$	
	$\frac{(N-1)(N-2)(N-3)}{N^3}$				1			

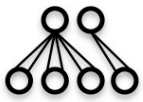
Pattern	P{Pattern}	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}	p_{37}
	$\frac{1}{N^3}$	$\frac{25}{216}$					$\frac{65}{108}$	$\frac{61}{216}$
	$\frac{N-1}{N^3}$			$\frac{25}{36}$			$\frac{5}{18}$	$\frac{1}{36}$

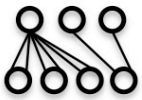
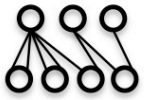
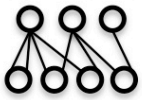
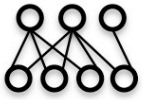
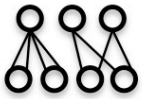
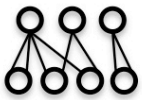
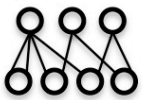
Pattern	P{Pattern}	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}	p_{37}
	$4 \bullet \frac{N-1}{N^3}$	$\frac{5}{216}$	$\frac{5}{18}$	$\frac{5}{36}$			$\frac{107}{216}$	$\frac{7}{108}$
	$2 \bullet \frac{N-1}{N^3}$	$\frac{1}{108}$	$\frac{2}{9}$	$\frac{11}{36}$			$\frac{23}{54}$	$\frac{1}{27}$
	$2 \bullet \frac{(N-1)(N-2)}{N^3}$			$\frac{5}{18}$	$\frac{5}{9}$		$\frac{1}{6}$	
	$4 \bullet \frac{(N-1)(N-2)}{N^3}$	$\frac{1}{216}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{5}{9}$		$\frac{13}{54}$	$\frac{1}{216}$
	$\frac{(N-1)(N-2)(N-3)}{N^3}$			$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$		

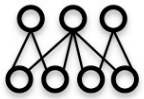
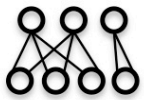
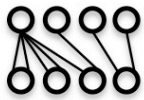
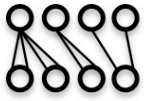
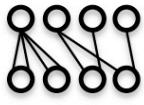
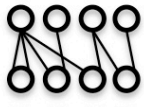
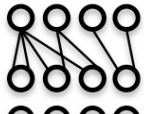
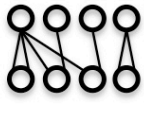
Pattern	P{Pattern}	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}	p_{46}	p_{47}
	$\frac{1}{N^5}$	$\frac{5}{48}$					$\frac{7}{12}$	$\frac{5}{16}$
	$6 \bullet \frac{N-1}{N^5}$	$\frac{1}{24}$	$\frac{2}{9}$	$\frac{1}{24}$			$\frac{79}{144}$	$\frac{7}{48}$
	$3 \bullet \frac{N-1}{N^5}$		$\frac{5}{18}$	$\frac{5}{24}$			$\frac{17}{36}$	$\frac{1}{24}$
	$12 \bullet \frac{N-1}{N^5}$	$\frac{1}{64}$	$\frac{5}{24}$	$\frac{7}{32}$			$\frac{23}{48}$	$\frac{5}{64}$
	$6 \bullet \frac{N-1}{N^5}$		$\frac{7}{36}$	$\frac{1}{3}$			$\frac{31}{72}$	$\frac{1}{24}$
	$4 \bullet \frac{N-1}{N^5}$	$\frac{1}{96}$	$\frac{5}{24}$	$\frac{1}{4}$			$\frac{15}{32}$	$\frac{1}{16}$
	$3 \bullet \frac{(N-1)(N-2)}{N^5}$		$\frac{5}{18}$	$\frac{1}{12}$	$\frac{1}{6}$		$\frac{4}{9}$	$\frac{1}{36}$

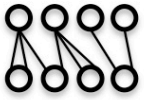
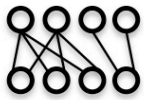
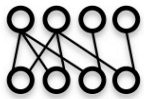
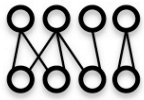
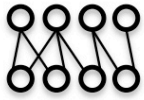
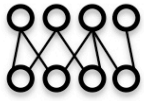
Pattern	P{Pattern}	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}	p_{46}	p_{47}
	$12 \bullet \frac{(N-1)(N-2)}{N^5}$	$\frac{1}{64}$	$\frac{13}{72}$	$\frac{1}{24}$	$\frac{7}{24}$		$\frac{59}{144}$	$\frac{35}{576}$
	$24 \bullet \frac{(N-1)(N-2)}{N^5}$		$\frac{1}{8}$	$\frac{5}{32}$	$\frac{3}{8}$		$\frac{47}{144}$	$\frac{5}{288}$
	$12 \bullet \frac{(N-1)(N-2)}{N^5}$		$\frac{7}{72}$	$\frac{5}{48}$	$\frac{1}{2}$		$\frac{41}{144}$	$\frac{1}{72}$
	$24 \bullet \frac{(N-1)(N-2)}{N^5}$	$\frac{1}{192}$	$\frac{19}{144}$	$\frac{11}{96}$	$\frac{3}{8}$		$\frac{199}{576}$	$\frac{1}{36}$
	$6 \bullet \frac{(N-1)(N-2)}{N^5}$		$\frac{1}{18}$	$\frac{5}{48}$	$\frac{7}{12}$		$\frac{1}{4}$	$\frac{1}{144}$
	$1 \bullet \frac{(N-1)(N-2)}{N^5}$				$\frac{5}{6}$		$\frac{1}{6}$	
	$8 \bullet \frac{(N-1)(N-2)}{N^5}$		$\frac{1}{12}$	$\frac{5}{32}$	$\frac{11}{24}$		$\frac{7}{24}$	$\frac{1}{96}$
	$12 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$		$\frac{7}{72}$	$\frac{1}{24}$	$\frac{1}{2}$	$\frac{1}{9}$	$\frac{35}{144}$	$\frac{1}{144}$
	$8 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$	$\frac{1}{192}$	$\frac{5}{48}$	$\frac{1}{32}$	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{49}{192}$	$\frac{1}{48}$
	$3 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$				$\frac{2}{3}$	$\frac{2}{9}$	$\frac{1}{9}$	
	$12 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$		$\frac{1}{36}$	$\frac{5}{96}$	$\frac{13}{24}$	$\frac{2}{9}$	$\frac{11}{72}$	$\frac{1}{288}$
	$6 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$		$\frac{1}{18}$	$\frac{1}{24}$	$\frac{7}{12}$	$\frac{1}{9}$	$\frac{5}{24}$	
	$24 \bullet \frac{(N-1)(N-2)(N-3)}{N^5}$		$\frac{1}{18}$	$\frac{7}{96}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{29}{144}$	$\frac{1}{288}$

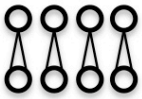
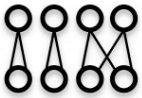
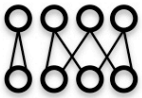
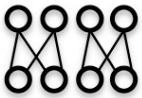
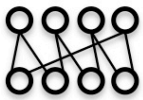
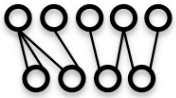
Pattern	P{Pattern}	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}	p_{46}	p_{47}
	$3 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^5}$				$\frac{1}{2}$	$\frac{4}{9}$	$\frac{1}{18}$	
	$12 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^5}$		$\frac{1}{36}$	$\frac{1}{48}$	$\frac{11}{24}$	$\frac{7}{18}$	$\frac{5}{48}$	
	$\frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^5}$				$\frac{1}{3}$	$\frac{2}{3}$		

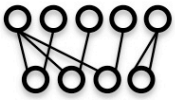
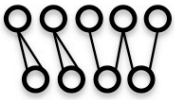
Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$\frac{1}{N^7}$	$\frac{3}{32}$					$\frac{9}{16}$	$\frac{11}{32}$
	$8 \bullet \frac{N-1}{N^7}$	$\frac{3}{64}$	$\frac{3}{16}$				$\frac{9}{16}$	$\frac{13}{64}$
	$4 \bullet \frac{N-1}{N^7}$		$\frac{3}{8}$				$\frac{9}{16}$	$\frac{1}{16}$
	$24 \bullet \frac{N-1}{N^7}$	$\frac{3}{128}$	$\frac{3}{16}$	$\frac{9}{64}$			$\frac{33}{64}$	$\frac{17}{128}$
	$24 \bullet \frac{N-1}{N^7}$		$\frac{3}{16}$	$\frac{9}{32}$			$\frac{15}{32}$	$\frac{1}{16}$
	$32 \bullet \frac{N-1}{N^7}$	$\frac{3}{256}$	$\frac{3}{16}$	$\frac{27}{128}$			$\frac{63}{128}$	$\frac{25}{256}$
	$3 \bullet \frac{N-1}{N^7}$			$\frac{9}{16}$			$\frac{3}{8}$	$\frac{1}{16}$
	$24 \bullet \frac{N-1}{N^7}$		$\frac{3}{16}$	$\frac{9}{32}$			$\frac{15}{32}$	$\frac{1}{16}$
	$8 \bullet \frac{N-1}{N^7}$	$\frac{3}{256}$	$\frac{3}{16}$	$\frac{27}{128}$			$\frac{63}{128}$	$\frac{25}{256}$

Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$4 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{8}$				$\frac{9}{16}$	$\frac{1}{16}$
	$24 \cdot \frac{(N-1)(N-2)}{N^7}$	$\frac{3}{128}$	$\frac{3}{16}$		$\frac{3}{16}$		$\frac{31}{64}$	$\frac{15}{128}$
	$24 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{16}$		$\frac{3}{8}$		$\frac{13}{32}$	$\frac{1}{32}$
	$48 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{16}$	$\frac{9}{64}$	$\frac{3}{16}$		$\frac{7}{16}$	$\frac{3}{64}$
	$96 \cdot \frac{(N-1)(N-2)}{N^7}$	$\frac{3}{256}$	$\frac{9}{64}$	$\frac{9}{128}$	$\frac{9}{32}$		$\frac{27}{64}$	$\frac{19}{256}$
	$24 \cdot \frac{(N-1)(N-2)}{N^7}$			$\frac{9}{32}$	$\frac{3}{8}$		$\frac{5}{16}$	$\frac{1}{32}$
	$96 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{32}$	$\frac{9}{64}$	$\frac{3}{8}$		$\frac{23}{64}$	$\frac{1}{32}$
	$96 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{9}{64}$	$\frac{9}{64}$	$\frac{9}{32}$		$\frac{51}{128}$	$\frac{5}{128}$
	$64 \cdot \frac{(N-1)(N-2)}{N^7}$	$\frac{3}{512}$	$\frac{9}{64}$	$\frac{27}{256}$	$\frac{9}{32}$		$\frac{105}{256}$	$\frac{29}{512}$
	$6 \cdot \frac{(N-1)(N-2)}{N^7}$				$\frac{3}{4}$		$\frac{1}{4}$	
	$12 \cdot \frac{(N-1)(N-2)}{N^7}$			$\frac{9}{32}$	$\frac{3}{8}$		$\frac{5}{16}$	$\frac{1}{32}$
	$48 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{32}$		$\frac{9}{16}$		$\frac{21}{64}$	$\frac{1}{64}$
	$96 \cdot \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{32}$	$\frac{9}{64}$	$\frac{3}{8}$		$\frac{23}{64}$	$\frac{1}{32}$

Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$48 \bullet \frac{(N-1)(N-2)}{N^7}$	$\frac{3}{512}$	$\frac{3}{32}$	$\frac{27}{256}$	$\frac{3}{8}$		$\frac{95}{256}$	$\frac{25}{512}$
	$24 \bullet \frac{(N-1)(N-2)}{N^7}$				$\frac{3}{4}$		$\frac{1}{4}$	
	$48 \bullet \frac{(N-1)(N-2)}{N^7}$			$\frac{9}{64}$	$\frac{9}{16}$		$\frac{9}{32}$	$\frac{1}{64}$
	$96 \bullet \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{64}$	$\frac{9}{64}$	$\frac{15}{32}$		$\frac{41}{128}$	$\frac{3}{128}$
	$16 \bullet \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{32}$		$\frac{9}{16}$		$\frac{21}{64}$	$\frac{1}{64}$
	$96 \bullet \frac{(N-1)(N-2)}{N^7}$		$\frac{3}{32}$	$\frac{9}{64}$	$\frac{3}{8}$		$\frac{23}{64}$	$\frac{1}{32}$
	$24 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{16}$		$\frac{3}{8}$		$\frac{13}{32}$	$\frac{1}{32}$
	$32 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$	$\frac{3}{256}$	$\frac{9}{64}$		$\frac{9}{32}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{17}{256}$
	$12 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{3}{4}$		$\frac{1}{4}$	
	$24 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{64}$	$\frac{9}{16}$		$\frac{9}{32}$	$\frac{1}{64}$
	$192 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{32}$	$\frac{9}{128}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{5}{16}$	$\frac{3}{128}$
	$48 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{32}$		$\frac{9}{16}$		$\frac{21}{64}$	$\frac{1}{64}$
	$48 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{32}$		$\frac{3}{8}$	$\frac{1}{4}$	$\frac{17}{64}$	$\frac{1}{64}$

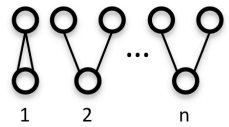
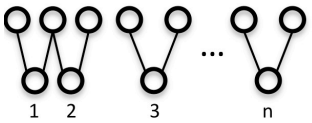
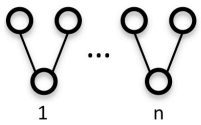
Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$96 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$	$\frac{3}{512}$	$\frac{3}{32}$	$\frac{9}{256}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{83}{256}$	$\frac{21}{512}$
	$24 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{3}{4}$		$\frac{1}{4}$	
	$48 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{64}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{7}{32}$	$\frac{1}{64}$
	$96 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{64}$	$\frac{9}{128}$	$\frac{15}{32}$	$\frac{1}{8}$	$\frac{35}{128}$	$\frac{1}{64}$
	$16 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{32}$		$\frac{9}{16}$		$\frac{21}{64}$	$\frac{1}{64}$
	$96 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{32}$	$\frac{9}{128}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{5}{16}$	$\frac{3}{128}$
	$24 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$96 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{9}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	
	$48 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{64}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{7}{32}$	$\frac{1}{64}$
	$192 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{128}$	$\frac{9}{16}$	$\frac{1}{8}$	$\frac{15}{64}$	$\frac{1}{128}$
	$96 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{64}$		$\frac{15}{32}$	$\frac{1}{4}$	$\frac{29}{128}$	$\frac{1}{128}$
	$192 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{64}$	$\frac{9}{128}$	$\frac{15}{32}$	$\frac{1}{8}$	$\frac{35}{128}$	$\frac{1}{64}$
	$192 \cdot \frac{(N-1)(N-2)(N-3)}{N^7}$		$\frac{3}{64}$	$\frac{9}{128}$	$\frac{15}{32}$	$\frac{1}{8}$	$\frac{35}{128}$	$\frac{1}{64}$

Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$1 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$					1		
	$12 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$32 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$				$\frac{9}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	
	$12 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{64}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{7}{32}$	$\frac{1}{64}$
	$48 \bullet \frac{(N-1)(N-2)(N-3)}{N^7}$			$\frac{9}{128}$	$\frac{9}{16}$	$\frac{1}{8}$	$\frac{15}{64}$	$\frac{1}{128}$
	$6 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{3}{4}$		$\frac{1}{4}$	
	$48 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$		$\frac{3}{32}$		$\frac{3}{8}$	$\frac{1}{4}$	$\frac{17}{64}$	$\frac{1}{64}$
	$16 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$	$\frac{3}{512}$	$\frac{3}{32}$		$\frac{9}{32}$	$\frac{5}{16}$	$\frac{69}{256}$	$\frac{19}{512}$
	$48 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$96 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{9}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	
	$96 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$			$\frac{9}{128}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{11}{64}$	$\frac{1}{128}$
	$96 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$		$\frac{3}{64}$		$\frac{15}{32}$	$\frac{1}{4}$	$\frac{29}{128}$	$\frac{1}{128}$
	$192 \bullet \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$		$\frac{3}{64}$	$\frac{9}{256}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{7}{32}$	$\frac{3}{256}$

Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$32 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$		$\frac{3}{64}$		$\frac{9}{32}$	$\frac{1}{2}$	$\frac{21}{128}$	$\frac{1}{128}$
	$4 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$					1		
	$24 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{3}{16}$	$\frac{3}{4}$	$\frac{1}{16}$	
	$24 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$96 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$48 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$			$\frac{9}{128}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{11}{64}$	$\frac{1}{128}$
	$32 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$				$\frac{9}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	
	$192 \cdot \frac{(N-1)(N-2)(N-3)(N-4)}{N^7}$			$\frac{9}{256}$	$\frac{15}{32}$	$\frac{5}{16}$	$\frac{23}{128}$	$\frac{1}{256}$
	$6 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$					1		
	$48 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$				$\frac{3}{16}$	$\frac{3}{4}$	$\frac{1}{16}$	
	$12 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$48 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$96 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$			$\frac{9}{256}$	$\frac{9}{32}$	$\frac{9}{16}$	$\frac{15}{128}$	$\frac{1}{256}$
	$24 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$				$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	
	$32 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)}{N^7}$		$\frac{3}{64}$		$\frac{9}{32}$	$\frac{1}{2}$	$\frac{21}{128}$	$\frac{1}{128}$

Pattern	P{Pattern}	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}	p_{56}	p_{57}
	$4 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)(N-6)}{N^7}$					1		
	$24 \cdot \frac{(N-1)(N-2)(N-3)(N-4)(N-5)(N-6)}{N^7}$				$\frac{3}{16}$	$\frac{3}{4}$	$\frac{1}{16}$	
	$\frac{(N-1)(N-2)(N-3)(N-4)(N-5)(N-6)(N-7)}{N^7}$					1		

Table S4 Patterns of ancestry that are *O*(1) and *O*(1/*N*) used in constructing the coalescent process for an arbitrary sample size *n*.

Pattern	P{Pattern}	Number of Parents	P{Coalescence|Pattern}
	$n\frac{1}{N}+O\left(\frac{1}{N^2}\right)$	$2n-1$	0
	$\left[\binom{2n}{2}-n\right]\frac{1}{N}+O\left(\frac{1}{N^2}\right)$	$2n-1$	$\frac{1}{16}$
	$1-\binom{2n}{2}\frac{1}{N}+O\left(\frac{1}{N^2}\right)$	$2n$	0

File S1

Extending Coalescent Theory to Self-Fertilizing Autotetraploids

For a population of autotetraploids that self-fertilize with probability s, the single-generation transition matrix for the ancestral process is given by

$$\mathbf{P} = \begin{bmatrix} \frac{1}{3}+\frac{1}{2}s & \frac{2}{3}(1-s) & \frac{s}{6} \\ \frac{3}{4N} & 1-\frac{1}{N} & \frac{1}{4N} \\ 0 & 0 & 1 \end{bmatrix}.$$

Using the result of Möhle 1998 and collecting the "fast" events, which occur on the timescale of single generations, into matrix **F** and the "slow" events, which occur on the timescale of N generations, into matrix **S**, we obtain

$$\mathbf{F} = \begin{bmatrix} \frac{1}{3}+\frac{s}{2} & \frac{2(1-s)}{3} & \frac{s}{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{3}{4} & -1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

in the limit as *N* tends to infinity. Here, $\mathbf{F} = \lim_{N\to\infty} \mathbf{P}$ and $\mathbf{S} = \lim_{N\to\infty} N(\mathbf{P}-\mathbf{F})$. If the limit $\mathbf{E} = \lim_{t\to\infty} \mathbf{F}^t$ exists, the continuous-time approximation to this ancestral process with partial selfing is characterized by matrices

$$\mathbf{E} = \begin{bmatrix} 0 & \frac{4-4s}{4-3s} & \frac{s}{4-3s} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{G} = \begin{bmatrix} 0 & -\frac{12s^2-22s+10}{9s^2-24s+16} & \frac{4-4s}{9s^2-24s+16} \\ 0 & -\frac{1}{4-3s} & \frac{1}{4-3s} \\ 0 & 0 & 0 \end{bmatrix}$$

where $\mathbf{G} = \mathbf{ESE}$ (Möhle 1998).

Note on extending double reduction to arbitrary *n*

If four lineages are within an individual and double reduction is possible, three types of events may occur for each pair of lineages in the instantaneous adjustment of the sample: (1) two lineages come from the same gamete and coalesce in the immediately previous generation via double reduction with probability α^2, (2) two lineages come from the same gamete, do not coalesce in the immediately previous generation, but coalesce in later generations via double reduction with probability $(1-\alpha)\frac{\alpha}{\alpha+2}$, or (3) two lineages get separated into distinct individuals before any coalescence event with probability $(1-\alpha)\left(\frac{2}{\alpha+2}\right)$. Putting these probabilities together, if four lineages are within an individual, a double coalescence event occurs in the instantaneous adjustment of the sample with probability $\alpha^2 + 2\alpha(1-\alpha)\frac{\alpha}{\alpha+2} + (1-\alpha)^2\left(\frac{\alpha}{\alpha+2}\right)^2$, a single coalescence event occurs with probability $2\left(\alpha(1-\alpha)\left(\frac{2}{\alpha+2}\right)\right) + 2\left((1-\alpha)^2\left(\frac{\alpha}{\alpha+2}\right)\left(\frac{2}{\alpha+2}\right)\right)$, and no coalescence events occur with probability $(1-\alpha)^2\left(\frac{2}{\alpha+2}\right)^2$. Thus, the number of instantaneous coalescence events for *m* individuals that contain four lineages is Multinomial with length *m* and probability vector **p** containing the three probabilities described above.