

# The third moments of the site frequency spectrum

A. Klassmann<sup>a,\*</sup>, L. Ferretti<sup>b</sup>

<sup>a</sup>*Institut für Genetik, Universität zu Köln, 50674 Köln, Germany*

<sup>b</sup>*The Pirbright Institute, Woking, United Kingdom*

---

## Abstract

The analysis of patterns of segregating (i.e. polymorphic) sites in aligned sequences is routine in population genetics. Quantities of interest include the total number of segregating sites and the number of sites with mutations of different frequencies, the so-called *site frequency spectrum*. For neutrally evolving sequences, some classical results are available, including the expected value and variance of the spectrum in the Kingman coalescent model without recombination as calculated by FU (1995).

In this work, we use similar techniques to compute the third moments of the site frequency spectrum without recombination. We also account for the linkage pattern of mutations, yielding the full haplotype spectrum of three polymorphic sites. Based on these results, we derive analytical results for the bias of Tajima's  $D$  and other neutrality tests.

As an application, we obtain the second moments of the spectrum of linked sites, which is related to the neutral spectrum of chromosomal inversions and other structural variants. These moments can be used for the normalisation of new neutrality tests relying on these spectra.

*Keywords:* Single nucleotide polymorphisms; infinite-sites model; site frequency spectrum; coalescent approximation; nested mutations; skewness.

---

\*Corresponding author

## 1. Introduction

Statistics based on polymorphic loci are key to estimate relevant quantities in population genetics, such as the rescaled mutation rate  $\theta$ . One common approach is to group variants together that appear with the same frequency in a sample and count the elements of each such group. The resulting summary statistic is called the *site frequency spectrum*.

The frequency spectrum is one of the most relevant statistics for population genetics. It can be used to infer evolutionary parameters such as mutation and recombination rate, past population history, demography and selection (HUDSON, 1983; NIELSEN *et al.*, 2005; HEIN *et al.*, 2004). Often, the variants are biallelic SNPs that can be “polarized”, e.g. for each site it is possible to say which allele is ancestral and which one is derived. This is the case for sequences with low mutation rate per base and for which an outgroup sequence is available. In what follows, we will consider exclusively this situation and assume that the evolution of these sequences can be modelled by a standard neutral Wright-Fisher model of constant population size.

WATTERSON (1975) credits FISHER (1930) with the first derivation (for a special case) of the first moments of the frequency spectrum. Their statement for the continuous case can be found in (EWENS, 1979), where it follows from results of diffusion theory (KIMURA, 1964). WATTERSON (1975) himself derived the first and second moments for the sum over all classes of the frequency spectrum, i.e. the number of segregating sites, using the technique of “moment estimators”. The full distribution of this quantity was shown by TAVARÉ (1984). The first and second moments for combinations of some components of the spectrum were later computed by TAJIMA (1989) using coalescent theory (KINGMAN, 1982) and combinatorics, while FU (1995) completed this approach for the full frequency spectrum. A major application of his formulas is the normalisation of a class of neutrality tests such as Tajimas’s  $D$  (TAJIMA, 1989), as described by ACHAZ (2009). Recently, HUDSON (2015) has given another proof of the first moments. As far as we know, higher moments of the spectrum have never been computed.

31 Asymptotic results for the distribution of the spectrum have been obtained  
32 by DAHMER and KERSTING (2015). Their approach shows that for sample  
33 size  $n \rightarrow \infty$  and large  $\theta$  (i.e. ignoring mutational Poisson noise), the first  $k$   
34 components of the spectrum converge in distribution to i.i.d normal variables  
35 with mean  $\theta/i$  and variance  $\theta^2 \ln(n)/n$ . However, their method applies only to  
36 allele counts which are much smaller than the sample size, hence it does not  
37 provide information on the full frequency spectrum in finite samples.

38 In this article we derive exact expressions for the third moments of the fre-  
39 quency spectrum. We use notation and approach of FU (1995), with some  
40 technical modifications in order to keep the number of different cases manage-  
41 able. We derive independently by the approach of WATTERSON (1975) the third  
42 moment of the number of segregating sites and show the consistency of the two  
43 approaches. An immediate corollary of the third moments is the expected fre-  
44 quency spectrum for three linked segregating sites, which fully characterises the  
45 expected haplotype structure for triplets of sites.

46 We present two applications. The first one concerns the bias of neutrality  
47 tests. Several neutrality tests based on the frequency spectrum, like Tajima's  
48  $D$ , should ideally have an expected value of zero, yet they don't. For the first  
49 time, we obtain general expressions for the bias of these neutrality tests as a  
50 function of mutation rate and sample size.

51 Finally, we derive the variance of the frequency spectra of nested and disjoint  
52 mutations at sites linked to a focal mutation. These spectra are equivalent to  
53 the spectrum of neutrally evolving chromosomal inversions (FERRETTI *et al.*,  
54 2017). Moreover, they represent the basis for the derivation of the spectra of  
55 other structural variants. With these results, it is possible to obtain the proper  
56 normalisation for new Tajima's  $D$ -like tests relying on the spectrum of linked  
57 mutations, e.g. neutrality tests for chromosomal introgressions or inversions.

## 58 **2. Results**

59 As is common practise in coalescent theory, we define  $\theta$  as the population-  
60 scaled mutation rate per sequence, i.e.  $\theta = 2pN_e\mu L$  where  $p$  is the ploidy,  $N_e$  is

61 the effective population size,  $\mu$  is the mutation rate per generation per bp and  $L$   
 62 is the length of the sequence in base pairs. We refer to the number of mutations  
 63 of size  $i$  in a sample of  $n$  sequences (i.e. the frequency spectrum) as  $\xi_i$ .

64 The model that we consider is the Kingman coalescent, with an infinite-sites  
 65 model of mutations. We assume no recombination, i.e. complete linkage among  
 66 sites.

### 67 2.1. The third moments of the frequency spectrum

68 Our main result is an analytical expression for the third moments of the  
 69 frequency spectrum.

70

#### Theorem 1.

$$E[\xi_h \xi_i \xi_j] = \delta_{h=i=j} \tau_i \theta + (\delta_{h=i} \tau_{ij} + \delta_{i=j} \tau_{hj} + \delta_{j=h} \tau_{hi}) \theta^2 + \tau_{hij} \theta^3 \quad (1)$$

for  $1 \leq h, i, j < n$ . The functions  $\tau$  are:

$$\tau_i = \frac{1}{i}, \quad (2)$$

$$\tau_{ij} = t_a(i, j) + t_a(j, i) + t_b(i, j) + t_b(j, i) \quad (3)$$

with

$$t_a(i, j) = \begin{cases} \frac{1}{2} (\beta_n(j) - \beta_n(j+1)) & \text{if } j < i \\ \frac{1}{2} \beta_n(j) & \text{if } j = i \end{cases} \quad (4)$$

$$t_b(i, j) = \begin{cases} \frac{1}{ij} - \frac{1}{i(i+j)} - \frac{1}{2} (\beta_n(j) - \beta_n(j+1)) & \text{if } i+j < n \\ \alpha_n(j) - \frac{1}{2} \beta_n(j) & \text{if } i+j = n, \end{cases}$$

and<sup>1</sup>

$$\tau_{hij} = \sum_{\text{Permutations}(h,i,j)} t_{aa}(h, i, j) + t_{ab}(h, i, j) + t_{ba}(h, i, j) + t_{bb}(h, i, j) \quad (5)$$

---

<sup>1</sup> $\sum_{\text{Perm.}(h,i,j)} f(h, i, j) = f(h, i, j) + f(i, j, h) + f(j, h, i) + f(h, j, i) + f(i, h, j) + f(j, i, h)$

with

$$\begin{aligned}
 t_{aa}(h, i, j) &= \begin{cases} \beta_n^5(i, j) & \text{if } j < i \text{ and } i < h \\ \beta_n^4(i, j) & \text{if } j < i \text{ and } i = h \\ \zeta_n(j) & \text{if } j = i \text{ and } i < h \\ \delta_n(j) - \alpha_n(j) + \frac{1}{2}\beta_n(j) & \text{if } j = i \text{ and } i = h \end{cases} \\
 t_{ab}(h, i, j) &= \begin{cases} \lambda_n(h - i, i, j) & \text{if } i + j < h \\ \beta_n^6(i, j) & \text{if } i + j = h \end{cases} \\
 t_{ba}(h, i, j) &= \begin{cases} \lambda_n(n - i, j, h) - \lambda_n(i, h, j) + \beta_n^2(h, j) - \beta_n^2(h, j + 1) - \beta_n^5(i, j) & \text{if } j < i \text{ and } h + i < n \\ \alpha_n^3(h, j) - \beta_n^3(h, j) + \alpha_n^4(n - h, j) - \beta_n^4(n - h, j) + \beta_n^6(j, h) & \text{if } j < i \text{ and } h + i = n \\ \beta_n^2(h, j) + \beta_n^3(j, h) - \beta_n^6(h, j) - \zeta_n(j) & \text{if } j = i \text{ and } h + i < n \\ \frac{1}{2} (\alpha_n(j) - \frac{1}{2}\beta_n(j) + \alpha_n(n - j) - \frac{1}{2}\beta_n(n - j)) + \gamma_n(j) & \text{if } j = i \text{ and } h + i = n \end{cases} \\
 t_{bb}(h, i, j) &= \begin{cases} \frac{1}{(h+i+j)(h+i)h} + \frac{1}{ij(h+i)} - \frac{1}{ih(i+j)} & \\ -\beta_n^2(i, j) + \beta_n^2(i, j + 1) - \lambda_n(n - h - i, i, j) & \text{if } h + i + j < n \\ \alpha_n^2(i, j) - \beta_n^2(i, j) - \beta_n^6(i, j) & \text{if } h + i + j = n, \end{cases} \tag{6}
 \end{aligned}$$

<sup>71</sup> using the following auxiliary functions (notation with upper indices):

$$\begin{aligned}
 a_n &= \sum_{i=1}^{n-1} \frac{1}{i} \\
 \alpha_n(i) &= \frac{a_n - a_i}{n - i} & \alpha_n^2(i, j) &= \frac{1}{i} (\alpha_n(j) - \alpha_n(i + j)) \\
 \beta_n(i) &= \frac{2}{\binom{n-1}{i} i} \sum_{k=2}^n \frac{\binom{n-k}{i-1}}{k} & \beta_n^2(i, j) &= \frac{1}{2i} (\beta_n(j) - \beta_n(i + j)) \\
 \gamma_n(i) &= \frac{1}{\binom{n-1}{i} i} \sum_{k=2}^n \frac{\binom{n-k}{i-1}}{k^2} & \zeta_n(i) &= \frac{1}{n - i - 1} \left( 2\alpha_n(i) - \frac{1}{2}\beta_n(i) - 2\delta_n(i) \right) \\
 \delta_n(i) &= \frac{1}{\binom{n-1}{i} i} \sum_{k=2}^n \frac{\binom{n-k}{i-1}}{k(k-1)} a_{k+1} & & \tag{7}
 \end{aligned}$$

$$\begin{aligned}
\alpha_n^3(i, j) &= \sum_{k=3}^n \sum_{t=1}^{k-2} \frac{\binom{i-1}{t-1} \binom{n-i-j-1}{k-t-2}}{\binom{n-1}{k-1}} \frac{k-t}{k(k-1)} \alpha_k(t) \\
\alpha_n^4(i, j) &= \sum_{k=3}^n \sum_{t=2}^{k-1} \frac{\binom{i-j-1}{t-2} \binom{n-i-1}{k-t-1}}{\binom{n-1}{k-1}} \frac{t}{k(k-1)} \alpha_k(t) \\
\beta_n^3(i, j) &= \sum_{k=3}^n \sum_{t=1}^{k-2} \frac{\binom{i-1}{t-1} \binom{n-i-j-1}{k-t-2}}{\binom{n-1}{k-1}} \frac{k-t}{k(k-1)} \frac{\beta_k(t)}{2} \\
\beta_n^4(i, j) &= \sum_{k=3}^n \sum_{t=2}^{k-1} \frac{\binom{i-j-1}{t-2} \binom{n-i-1}{k-t-1}}{\binom{n-1}{k-1}} \frac{t}{k(k-1)} \frac{\beta_k(t)}{2} \\
\beta_n^5(i, j) &= \frac{1}{n-i-1} \left( \frac{\beta_n(j) - \beta_n(j+1)}{2} - 2\beta_n^4(i, j) \right) \\
g_n(i, j) &= \sum_{k=2}^n \sum_{t=1}^{k-1} \frac{\binom{i-1}{t-1} \binom{n-i-j}{k-t-1}}{\binom{n-1}{k-1}} \frac{1}{k(k-1)} \beta_k(t) \\
\beta_n^6(i, j) &= \frac{1}{n-i-j} \left( \alpha_n(j) - \frac{\beta_n(j)}{2} - g_n(i, j) \right) + \alpha_n^2(i, j) - \beta_n^2(i, j)
\end{aligned} \tag{8}$$

$$\lambda_n(h, i, j) = (h-j+1)\beta_n^6(i, j) - 2(h-j)\beta_n^6(i, j+1) + (h-j-1)\beta_n^6(i, j+2) \tag{9}$$

**Remark 1.** The coefficient for  $\theta$  is the well known result for the expectation of the frequency spectrum

$$E[\xi_i] = \tau_i \theta = \frac{\theta}{i}. \tag{10}$$

73 The terms  $\tau_{ij}$  are identical to the quadratic part of the second moments,

$$E[\xi_i \xi_j] = \delta_{i=j} \tau_i \theta + \tau_{ij} \theta^2, \tag{11}$$

74 computed by FU (1995):  $\tau_{ij} = \sigma_{ij} + \frac{1}{ij}$ , with  $\sigma_{ij}$  defined in eq. (2) and (3)  
75 therein.

76

**Remark 2.** FU (1995) showed in his eq. (34), that for  $\beta_n(i)$  exists a more compact form, namely

$$\beta_n(i) = \frac{2n}{(n-i+1)(n-i)} (a_{n+1} - a_i) - \frac{2}{n-i},$$

77 where the summation over  $k$  is hidden in the  $a_n$ . We do not have a similar form  
78 for the other expressions, however  $\beta_n^3$  and  $\beta_n^4$  can be expressed in terms of  $g_n$ ,  
79 too. Hence in a computational implementation the speed limiting factor are the  
80 double summations in  $\alpha_n^3(i, j)$ ,  $\alpha_n^4(i, j)$  and  $g_n(i, j)$ .

81

**Remark 3.** The sum over permutations simplifies the fractions in  $t_b$  resp.  $t_{bb}$

$$\sum_{\text{Permutations}(i,j)} \left( \frac{1}{ij} - \frac{1}{i(i+j)} \right) = \frac{1}{ij} \quad (12)$$

$$\sum_{\text{Permutations}(h,i,j)} \left( \frac{1}{(h+i+j)(h+i)h} + \frac{1}{ij(h+i)} - \frac{1}{ih(h+j)} \right) = \frac{1}{hij}. \quad (13)$$

**Remark 4.** The central third moments can be obtained by

$$\begin{aligned} \mu_3[\xi_h, \xi_i, \xi_j] &= E[(\xi_h - E[\xi_h])(\xi_i - E[\xi_i])(\xi_j - E[\xi_j])] \\ &= E[\xi_h \xi_i \xi_j] - E[\xi_h]E[\xi_i \xi_j] - E[\xi_i]E[\xi_h \xi_j] - E[\xi_j]E[\xi_h \xi_i] + 2E[\xi_h]E[\xi_i]E[\xi_j]. \end{aligned} \quad (14)$$

**Remark 5.** For the folded spectrum

$$\eta_i = \frac{\xi_i + \xi_{n-i}}{1 + \delta_{i=n-i}}$$

the corresponding third moments can be computed in a similar way as the second moments (eq. (9) in FU (1995))

$$\begin{aligned} E[\eta_h \eta_i \eta_j] &= (E[\xi_h \xi_i \xi_j] + E[\xi_h \xi_i \xi_{n-j}] + E[\xi_h \xi_{n-i} \xi_j] + E[\xi_h \xi_{n-i} \xi_{n-j}] \\ &\quad + E[\xi_{n-h} \xi_i \xi_j] + E[\xi_{n-h} \xi_i \xi_{n-j}] + E[\xi_{n-h} \xi_{n-i} \xi_j] + E[\xi_{n-h} \xi_{n-i} \xi_{n-j}]) \\ &\quad \cdot \frac{1}{(1 + \delta_{h=n-h})(1 + \delta_{i=n-i})(1 + \delta_{j=n-j})}. \end{aligned} \quad (15)$$

## 82 2.2. The frequency spectrum of three linked sites

83 The components  $t_{aa}$ ,  $t_{ab}$ ,  $t_{ba}$  and  $t_{bb}$  correspond to different linkage patterns  
84 of three mutations (without recombination). The analogous pattern for two sites  
85 has been recently studied in FERRETTI *et al.* (2017). Pairs of linked mutations  
86 are *nested*, when one mutation is present only in sequences containing the other,  
87 or *disjoint*, if the mutations are present in different sets of sequences. The nested  
88 and disjoint components of the frequency spectrum for pairs of sites give a  
89 complete description of the haplotype structure of two sites (up to permutations  
90 of individuals and sites).

91 Following the derivation in FERRETTI *et al.* (2017), the frequency spectrum  
 92 for triplets of segregating sites is given by

$$E[\xi_{h,i,j}] = \begin{cases} E[\xi_h \xi_i \xi_j] & \text{for } h \neq i, h \neq j, i \neq j \\ E[\xi_h \xi_i (\xi_i - 1)]/2 = (E[\xi_h \xi_i^2] - E[\xi_h \xi_i])/2 & \text{for } i = j \neq h \\ E[\xi_h \xi_i (\xi_h - 1)]/2 = (E[\xi_h^2 \xi_i] - E[\xi_h \xi_i])/2 & \text{for } h = j \neq i \\ E[\xi_h \xi_j (\xi_h - 1)]/2 = (E[\xi_h^2 \xi_j] - E[\xi_h \xi_j])/2 & \text{for } h = i \neq j \\ E[\xi_h (\xi_h - 1)(\xi_h - 2)]/6 = (E[\xi_h^3] - 3E[\xi_h^2] + 2E[\xi_h])/6 & \text{for } h = i = j \end{cases} \quad (16)$$

93 noticing that three derived mutations in different sites can have four possible  
 94 relations:

- 95 • *fully nested*: the second mutation  $i$  is nested inside the first  $h$ , and the  
 96 third  $j$  is nested inside the second  $i$ .

97 This component corresponds to  $t_{aa}(h, i, j)$ .

- 98 • *disjoint within nested*: the second and third mutations  $i, j$  are disjoint,  
 99 but both are nested inside the first  $h$ .

100 This component corresponds to  $t_{ab}(h, i, j)$ .

- 101 • *nested within disjoint*: the mutations  $h, i$  are mutually disjoint, but mu-  
 102 tation  $j$  is nested inside  $i$  (and consequently  $j$  is disjoint to  $h$ , too).

103 This component corresponds to  $t_{ba}(h, i, j)$ .

- 104 • *fully disjoint*: all mutations  $h, i, j$  are mutually disjoint.

105 This component corresponds to  $t_{bb}(h, i, j)$ .

106 Therefore, the spectrum of three sites can be easily decomposed by separating  
 107 the components  $t_{aa}$ ,  $t_{ab}$ ,  $t_{ba}$  and  $t_{bb}$ . This spectrum is equivalent to a complete  
 108 characterization of the haplotype spectrum of three sites.



109 *2.3. Comparison with the third moments of the number of segregating sites*

110 We derive by the method of WATTERSON (1975) the third moments for the  
 111 number of segregating sites  $S = \sum_{i=1}^{n-1} \xi_i$ .

**Theorem 2.** *Writing  $H_{n,m} = \sum_{i=1}^n \frac{1}{i^m}$  for the  $n$ -th harmonic number of order  $m$ , the third moment (resp. central moment) of the number of segregating sites  $S$  for a sample of size  $n$  is:*

$$\begin{aligned} E[S^3] &= H_{n-1,1}\theta + 3(H_{n-1,1}^2 + H_{n-1,2})\theta^2 + (H_{n-1,1}^3 + 3H_{n-1,1}H_{n-1,2} + 2H_{n-1,3})\theta^3 \\ \mu_3[S] &= E[(S - E[S])^3] = H_{n-1,1}\theta + 3H_{n-1,2}\theta^2 + 2H_{n-1,3}\theta^3 \end{aligned}$$

(17)

Since

$$\sum_{h=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} E[\xi_h \xi_i \xi_j] = E[S^3]$$

and

$$\sum_{h=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \mu_3[\xi_h, \xi_i, \xi_j] = \mu_3[S],$$

112 following from Theorem 1 and Theorem 2, the corresponding coefficients for  $\theta$ ,  
 113  $\theta^2$  and  $\theta^3$  have to be the same. We give in the supplement an explicit proof of  
 114 the non-trivial identities of the coefficients for  $\theta^2$  and  $\theta^3$  stated as:

**Lemma 1.**

$$\begin{aligned} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \tau_{ij} &= H_{n-1,1}^2 + H_{n-1,2} \\ \sum_{h=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \tau_{hij} &= H_{n-1,1}^3 + 3H_{n-1,1}H_{n-1,2} + 2H_{n-1,3}. \end{aligned}$$

115 *2.4. Skewness and bias of Tajima's D and similar neutrality tests*

116 One of the applications of the frequency spectrum is to test if the observed  
 117 patterns in sequences are compatible with neutral evolutionary models. Several

118 neutrality tests fall into a general class that relies on normalized linear com-  
 119 binations of the frequency spectrum (ACHAZ (2009), FERRETTI *et al.* (2010)).  
 120 Their general form is

$$T_{\Omega} = \frac{\sum_{i=1}^{n-1} i\Omega_i\xi_i}{\sqrt{\text{Var}[\sum_{i=1}^{n-1} i\Omega_i\xi_i]}} \quad , \quad \sum_{i=1}^{n-1} \Omega_i = 0 \quad (18)$$

where the variance in the denominator

$$\text{Var} \left[ \sum_{i=1}^{n-1} i\Omega_i\xi_i \right] = \theta \sum_{i=1}^{n-1} i^2\Omega_i^2\tau_i + \theta^2 \sum_{i,j=1}^{n-1} ij\Omega_i\Omega_j \left( \tau_{ij} - \frac{1}{ij} \right)$$

121 is a linear combination of  $\theta$  and  $\theta^2$ . These two quantities, if unknown, are  
 122 usually estimated from  $S$  and  $S^2$  by the method of moments:  $\hat{\theta} = S/H_{n-1,1}$   
 123 and  $\hat{\theta}^2 = S(S-1)/(H_{n-1,1}^2 + H_{n-1,2})$ .

124 In this section, we explore the additional information that the third moments  
 125 of the spectrum reveal about the distribution of neutrality tests, in particular  
 126 about their skewness and bias.

127 [Table 1 about here.]

128 First, we consider the case of known  $\theta$ . It is well known the distributions of  
 129 neutrality tests based on the frequency spectrum such as Tajima's  $D$  (TAJIMA,  
 130 1989) tend to be skewed (HUDSON, 1991). These tests are normalized to mean  
 131 0 and variance 1 under the neutral coalescent with constant population size:  
 132  $E[T_{\Omega}] = 0$  and  $\text{Var}[T_{\Omega}] = 1$ . Consequently, the skewness  $\gamma = \mu_3/\sigma^3$  equals the  
 133 third moment of the test:

$$\gamma(T_{\Omega}) = E[T_{\Omega}^3] = \frac{\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} ijk\Omega_i\Omega_j\Omega_k \cdot E[\xi_i\xi_j\xi_k]}{\text{Var}[\sum_{i=1}^{n-1} i\Omega_i\xi_i]^{3/2}} \quad (19)$$

134 The weights  $\Omega_i$  for some commonly used neutrality tests are given in Table  
 135 1. Figure 1 shows, that analytical results and those from simulations with 'ms'  
 136 (HUDSON, 2002) agree well. However, when the parameter  $\theta$  has to be estimated  
 137 from the data, as it is usually the case, the denominator of the test is a function  
 138 of the estimator, contributing to the skewness. This has a relatively large effect,  
 139 but surprisingly for most considered values of  $\theta$  it reduces the skewness.

140

141 [Figure 1 about here.]

142 [Figure 2 about here.]

For  $\theta$  unknown and estimated from  $S$ , we can still make use of the third moments. In this case, we can compute an approximate result for the bias of the test. We apply the following formula for the Taylor expansion of moments of random variables<sup>2</sup>  $X, Y$  with  $E[X] = 0$  and  $Y > 0$

$$E\left[\frac{X}{\sqrt{Y}}\right] \approx -\frac{E[XY]}{2E[Y]^{3/2}}$$

143 and the fact that  $E[\sum_{k=1}^{n-1} k\Omega_k\xi_k] = 0$  to obtain the bias:

$$E[T_{\Omega}] \approx -\frac{\sum_{k=1}^{n-1} k\Omega_k \left[ E[\xi_k S] / a_n \sum_{i=1}^{n-1} i^2 \Omega_i^2 \tau_i + E[\xi_k S(S-1)] / (a_n^2 + b_n) \sum_{i,j=1}^{n-1} ij\Omega_i\Omega_j \left( \tau_{ij} - \frac{1}{ij} \right) \right]}{2 \left[ \theta \sum_{i=1}^{n-1} i^2 \Omega_i^2 \tau_i + \theta^2 \sum_{i,j=1}^{n-1} ij\Omega_i\Omega_j \left( \tau_{ij} - \frac{1}{ij} \right) \right]^{3/2}} \quad (20)$$

144 with  $E[\xi_k S] = \sum_{i=1}^{n-1} E[\xi_k \xi_i]$  resp.  $E[\xi_k S^2] = \sum_{i,j=1}^{n-1} E[\xi_k \xi_i \xi_j]$ .

145 The above equation gives a reasonably good estimate of the bias of neutrality  
146 tests (Figure 2), taking into account that eq. (20) represents only the first term  
147 of a bivariate Taylor expansion.

### 148 2.5. The variance of the frequency spectrum of linked sites

149 We will use the nomenclature introduced by SARGSYAN (2015) and expanded  
150 in FERRETTI *et al.* (2017). We call a certain mutation of interest *focal* and we  
151 refer to it as  $\phi$ . As above, further mutations that appear in at least one individ-  
152 ual together with it, are called *nested* while all others are called *disjoint*. Note,  
153 that the focal mutation  $\phi$  itself is not included into neither. More specifically,  
154 we refer to the number of mutations of size  $i$  that are nested with the focal  
155 mutation by  $\xi_{i,\phi}^N$  and to those that are disjoint by  $\xi_{i,\phi}^D$ . Evidently, the number  
156 of overall occurrences of mutations of size  $i$ , given  $\phi$ , is  $\xi_{i,\phi} = \xi_{i,\phi}^N + \xi_{i,\phi}^D$ .

<sup>2</sup>From the general expansion (e.g. VAN ERP and VAN GELDER (2007))

$$E\left[\frac{X}{\sqrt{Y}}\right] \approx \frac{E[X]}{\sqrt{E[Y]}} - \frac{E[XY] - E[X]E[Y]}{2E[Y]^{3/2}} + \frac{3E[X]\text{Var}[Y]}{8E[Y]^{5/2}}$$

157 We now condition on the focal mutation  $\phi$  being a mutation of any size  $h$  and  
 158 write:  $E[\xi_{i|h}^N] = E[\xi_{i,\phi}^N]/E[\xi_h]$  resp.  $E[\xi_{i|h}^D] = E[\xi_{i,\phi}^D]/E[\xi_h]$ . The expectation  
 159 values of nested mutations are a corollary of the second moments derived by  
 160 FU (1995); they are given in FERRETTI *et al.* (2017). The second moments of  
 161 two nested respective two disjoint mutations, or between one nested and one  
 162 disjoint mutation, are obtained directly from the above expressions of the third  
 163 moments:

**Corollary 1.**

$$\begin{aligned}
 E[\xi_{i|h}^N \xi_{j|h}^N] &= h \left( \delta_{i=j} t_a(h, i) \theta + \left( t_{ab}(h, i, j) + t_{ab}(h, j, i) + \sum_{\text{Permutations}(h, i, j)} t_{aa}(h, i, j) \right) \theta^2 \right) \\
 E[\xi_{i|h}^N \xi_{j|h}^D] &= h (t_{ab}(i, j, h) + t_{ab}(i, h, j) + t_{ba}(j, h, i) + t_{ba}(j, i, h)) \theta^2 \\
 E[\xi_{i|h}^D \xi_{j|h}^N] &= h (t_{ab}(j, i, h) + t_{ab}(j, h, i) + t_{ba}(i, h, j) + t_{ba}(i, j, h)) \theta^2 \\
 E[\xi_{i|h}^D \xi_{j|h}^D] &= h \left( \delta_{i=j} t_b(h, i) \theta + \left( t_{ba}(h, i, j) + t_{ba}(h, j, i) + \sum_{\text{Permutations}(h, i, j)} t_{bb}(h, i, j) \right) \theta^2 \right).
 \end{aligned}
 \tag{21}$$

## 164 2.6. Numerical results

165 In Figure 4 we compare the analytical results with the third moments from  
 166 coalescent simulations. We use “ms” (HUDSON, 2002) to generate samples and  
 167 from their frequency spectra we calculate estimates of the third moments. For  
 168 increasing sample size  $n$  the “off-diagonal” elements of the three-dimensional  
 169 array of third moments get increasingly small; that’s why the maximum rela-  
 170 tive error of the simulated data increases with  $n$ . The graphs show that with  
 171 increasing number of samples, the values from simulations converge to our an-  
 172 alytical results. However the convergence is extremely slow, indicating a large  
 173 variance of the third moments.

174 [Figure 3 about here.]

175 Figure 4 shows all third moments for a sample of size  $n = 5$ . As in the two-  
 176 dimensional case, the values of the diagonals (now in 3 dimensions) dominate.

177

178 In Figure 5 we compare the covariances of the standard frequency spectrum  
 179 with the covariances between nested and disjoint mutations of a conditional  
 180 spectrum. The spectra of nested, resp. disjoint, sites are still dominated by the  
 181 variances, while the correlation of “mirror sites” ( $\xi_i$  and  $\xi_{n-i}$  in the standard  
 182 spectrum), is lost. There is almost no correlation between nested and disjoint  
 183 sites.

184 [Figure 4 about here.]

185 [Figure 5 about here.]

186 *2.7. Comparison with asymptotic analytical results*

187 DAHMER and KERSTING (2015) showed the convergence of the distribution  
 188 of the components of the spectrum to centered and rescaled i.i.d. Gaussian  
 189 variables in the large  $n$  limit. More precisely, they state that for large  $\theta$ , i.e.  
 190 ignoring the Poisson noise, we have for fixed  $k$

$$\sqrt{\frac{n}{\ln(n)}} \left( \xi_1 - \theta, \xi_2 - \frac{\theta}{2}, \dots, \xi_k - \frac{\theta}{k} \right) \xrightarrow{n \rightarrow \infty} N(0, \theta^2 \cdot \mathbf{1}_{k \times k}). \quad (22)$$

One could naively assume, that this means that in the limit of large  $n$  the  $\xi_k$ s  
 could be treated as independent Gaussian random variables with mean  $\theta/k$  and  
 variance  $\theta^2 \ln(n)/n$ , leading to the approximation

$$\mathbb{E}[\xi_h \xi_i \xi_j]_{|\theta^3} = \tau_{hij} \theta^3 = \mathbb{E}[\xi_h] \mathbb{E}[\xi_i] \mathbb{E}[\xi_j] + \frac{\ln(n)}{n} (\delta_{h=i} \mathbb{E}[\xi_j] + \delta_{h=j} \mathbb{E}[\xi_i] + \delta_{i=j} \mathbb{E}[\xi_h]) \theta^2 + o\left(\frac{\ln(n)}{n}\right) \quad (23)$$

191 This is however incorrect. The distribution of each component of the spectrum  
 192  $\xi_k$  shows excesses of outliers and heavy tails (JANSON and KERSTING, 2011),  
 193 hence the convergence in distribution proved by Dahmer and Kersting does not  
 194 imply the scaling of the moments.

195 Figure 6 shows, that the asymptotics are of limited help for a particular  
 196 finite sample size  $n$ , since only moments for  $h, i, j \ll \frac{n}{2}$ , and only those with  
 197 at least two indices differing, seem to be approximated reasonably well.

198 [Figure 6 about here.]

199 **3. Methods**

200 *3.1. Proof of theorem 1*

201 *3.1.1. Separation of estimation*

202 A coalescent tree is constructed by two independent stochastic processes,  
203 namely its branching pattern (the topology) and the lengths of its branches (co-  
204 alescent times). The idea of FU (1995) is to decompose the tree into small parts,  
205 called *lines*, by cutting each branch along *states*, which are delineated by coa-  
206 lescent events. He first calculates all possible hierarchical relationships between  
207 those lines, thereby transforming a probabilistic problem into a combinatorial  
208 one. Second, he computes the estimated mutations on each line. This number  
209 is correlated between lines of the same state because of their shared lengths.  
210 The combined sum over the two quantities yields the desired second moments.  
211 We re-use method and notation of FU (1995) with appropriate extensions. A  
212 thorough explanation of the main ingredients of his proof, albeit with somewhat  
213 different notation, has been given in DURRETT (2008). An extended “reprint”  
214 of the more technical parts can be found in the supplement of our companion  
215 paper FERRETTI *et al.* (2017).

216 [Figure 7 about here.]

217 We define index variables  $\epsilon_{kl}(i)$ , that indicate if the line  $l$  of state  $k$  has  $i$   
218 descendants at state  $n$ , (e.g. they take the values 1 resp. 0). It follows that (cf.  
219 figure 7)

$$\xi_i = \sum_{k=2}^n \sum_{l=1}^k \epsilon_{kl}(i) \xi_{kl}. \quad (24)$$

220 In the following we use the fact, that the index  $l$  serves only to distinguish lines  
221 of the same state, but otherwise has no meaning, since all lines of the same  
222 state are equivalent. The indicator variables are idempotent ( $\epsilon_{kl}(i)^2 = \epsilon_{kl}(i)$ )  
223 and independent of the length (resp. mutation rate)  $\xi_{kl}$ . The expectation values  
224 of the indicator variables correspond to probabilities, which we will define in the  
225 following subsection.

226 3.1.2. Averaging over topologies

227

[Figure 8 about here.]

We split the computation of the expectation values of the indicator variables (which define the topology) into several cases, pictured in figure 8.

We recall, that the number of descendants of lines in the coalescent is equivalent to that of balls of a specific colour in a so-called *Pólya urn model* whose probability distribution is known and reviewed in e.g. GRIFFITHS and TAVARÉ (2003). We introduce the following notation:  $p_{k \rightarrow n}(t \rightarrow i)$  is the probability that  $t$  lines at state  $k$  have  $i$  descendants at state  $n$ . This probability is

$$p_{k \rightarrow n}(t \rightarrow i) = \frac{\binom{i-1}{t-1} \binom{n-i-1}{k-t-1}}{\binom{n-1}{k-1}}. \quad (25)$$

228 At this point it is helpful to define  $\binom{-1}{-1} = 1$ , while binomial coefficients con-  
 229 taining any other combination of one or two negative numbers are set to zero  
 230 (DURRETT, 2008). This makes it possible to subsume in the above and follow-  
 231 ing formulas the case that  $t = k$  lines of state  $k$  yield  $i = n$  lines at state  $n$   
 232 (which is true with probability 1). Later on, these special cases will be resolved  
 233 separately and none of the expressions in *Results* rely on this definition.

234

The probability that  $t$  and  $u$  (different) lines at state  $k$  have respectively  $i$  and  $j$  descendants at state  $n$  is

$$p_{k \rightarrow n}(t \rightarrow i, u \rightarrow j) = \frac{\binom{i-1}{t-1} \binom{j-1}{u-1} \binom{n-i-j-1}{k-t-u-1}}{\binom{n-1}{k-1}}. \quad (26)$$

And for three such (non-overlapping) sets of lines the probability yields

$$p_{k \rightarrow n}(s \rightarrow h, t \rightarrow i, u \rightarrow j) = \frac{\binom{h-1}{s-1} \binom{i-1}{t-1} \binom{j-1}{u-1} \binom{n-h-i-j-1}{k-s-t-u-1}}{\binom{n-1}{k-1}}. \quad (27)$$

Using this notation we can now state the probabilities for different configurations. We start with those derived by Fu: The probability, that one line at

state  $k$  has  $i$  descendants at state  $n$  is (FU, 1995, eq. (14))

$$\begin{aligned} p(k, i) &= p_{k \rightarrow n}(1 \rightarrow i) \\ &= \frac{\binom{n-i-1}{k-2}}{\binom{n-1}{k-1}}. \end{aligned} \quad (28)$$

The joint probability that one line at state  $k$  and one nested line at state  $k' \geq k$  have  $i$  respective  $j$  descendants at state  $n$  is (FU, 1995, eq. (18))

$$\begin{aligned} p_a(k, i; k', j) &= \sum_{t=1}^{k'-1} p_{k \rightarrow k'}(1 \rightarrow t) \frac{t}{k'} p_{k' \rightarrow n}(t-1 \rightarrow i-j, 1 \rightarrow j) \\ &= \sum_{t=1}^{k'-1} \frac{\binom{k'-t-1}{k-2}}{\binom{k'-1}{k-1}} \frac{t}{k'} \frac{\binom{i-j-1}{t-2} \binom{n-i-1}{k'-t-1}}{\binom{n-1}{k'-1}}. \end{aligned} \quad (29)$$

The joint probability that one line at state  $k$  and one disjoint (not nested) line at state  $k' \geq k$  have  $i$  resp.  $j$  descendants at state  $n$  is (FU, 1995, eq. (19) and (20))

$$\begin{aligned} p_b(k, i; k', j) &= \sum_{t=1}^{k'-1} p_{k \rightarrow k'}(1 \rightarrow t) \frac{k'-t}{k'} p_{k \rightarrow n}(t \rightarrow i, 1 \rightarrow j) \\ &= \sum_{t=1}^{k'-1} \frac{\binom{k'-t-1}{k-2}}{\binom{k'-1}{k-1}} \frac{k'-t}{k'} \frac{\binom{i-1}{t-1} \binom{n-i-j-1}{k'-t-2}}{\binom{n-1}{k'-1}} \end{aligned} \quad (30)$$

In the latter two cases the summation index  $t$  runs over the possible numbers of descendants that the line of state  $k$  may have at state  $k'$ . Since no single line can be ancestor of all  $k'$  lines, this number has an upper limit of  $k' - 1$ . There are more constraints on  $t$  as detailed by FU (1995) (e.g. a line from state  $k$  can have at most  $k' - k + 1$  descendants at state  $k'$ , hence only  $t \leq k' - k + 1$  is meaningful), however these are already implicit in the binomial coefficients.

Note, that Fu defined these equations only for the case  $k < k'$ . Using the special definition for the binomial coefficient, they include the case  $k = k'$  (DURRETT, 2008): if the lines are from the same state, then  $t = 1$  and we have  $p_a(k, i; k, j) = \delta_{i=j} \frac{1}{k} p(k, i)$  and  $p_b(k, i; k, j) = \frac{k-1}{k} \frac{\binom{n-i-j-1}{k-3}}{\binom{n-1}{k-1}}$ . These two equations correspond to eq. (14) and (15) of (FU, 1995).

Hence the probability, that a line at  $k$  and a line at state  $k \leq k'$  have  $i$  resp.  $j$



descendants at state  $n$  yields for  $2 \leq k \leq k' \leq n$ :

$$p(k, i; k', j) = p_a(k, i; k', j) + p_b(k, i; k', j). \quad (31)$$

235 Now we derive the probabilities involving three lines. These may be all of  
 236 the same state, of two different states or of three different states. We assume  
 237  $k \leq k' \leq k''$ . We take a single line at each state  $k$ ,  $k'$  and  $k''$  respectively and  
 238 subdivide along their possible relationships. We denote the lines  $l$ ,  $l'$  and  $l''$   
 239 respectively. The six cases are (compare figure 8):

- 240 •  $aa$ :  $l'$  is a descendant of  $l$  and  $l''$  is a descendant of  $l'$
- 241 •  $ab$ :  $l'$  and  $l''$  are both descendants of  $l$ , but  $l''$  is not a descendant of  $l'$
- 242 •  $ba^{(3)}$ :  $l'$  is a descendant of  $l$ , but  $l''$  is not
- 243 •  $ba^{(2)}$ :  $l''$  is a descendant of  $l$ , but  $l'$  is not
- 244 •  $ba^{(1)}$ :  $l''$  is a descendant of  $l'$ , but both are not descendants of  $l$
- 245 •  $bb$ : no line is a descendant of any of the other two lines

As before,  $t$  counts the number of descendants of line  $l$  at state  $k'$ .  $t_1$  denotes the number of descendants of  $l$  at state  $k''$ , without the descendants of  $l'$ .  $t_2$  finally counts the descendants of line  $l'$ . We present here only the first case, while all six cases are listed in the supplement.

$$\begin{aligned} & p_{aa}(k, h; k', i; k'', j) \\ &= \sum_{t=1}^{k'-1} \sum_{t_1=0}^{k''-2} \sum_{t_2=1}^{k''-t_1-1} p_{k \rightarrow k'}(1 \rightarrow t) \frac{t}{k'} p_{k' \rightarrow k''}(t-1 \rightarrow t_1, 1 \rightarrow t_2) \frac{t_2}{k''} p_{k'' \rightarrow n}(t_1 \rightarrow h-i, t_2-1 \rightarrow i-j, 1 \rightarrow j) \\ &= \sum_{t=1}^{k'-1} \sum_{t_1=0}^{k''-2} \sum_{t_2=1}^{k''-t_1-1} \frac{\binom{k'-t-1}{k-2}}{\binom{k'-1}{k-1}} \frac{t}{k'} \frac{\binom{t_1-1}{t-2} \binom{k''-t_1-t_2-1}{k'-t-1}}{\binom{k''-1}{k'-1}} \frac{t_2}{k''} \frac{\binom{h-i-1}{t_1-1} \binom{i-j-1}{t_2-2} \binom{n-h-1}{k''-t_1-t_2-1}}{\binom{n-1}{k''-1}}. \end{aligned} \quad (32)$$

Since the six cases cover all possible combinations, the total probability that three lines at state  $k$ ,  $k'$  and  $k''$  resp. (with  $k \leq k' \leq k''$ ) have  $h$ ,  $i$  and  $j$  resp.

descendants at state  $n$  is given by

$$\begin{aligned}
 p(k, h; k', i; k'', j) = & p_{aa}(k, h; k', i; k'', j) + p_{ab}(k, h; k', i; k'', j) + p_{ba}^{(3)}(k, h; k', i; k'', j) \\
 & + p_{ba}^{(2)}(k, h; k', i; k'', j) + p_{ba}^{(1)}(k, h; k', i; k'', j) + p_{bb}(k, h; k', i; k'', j).
 \end{aligned} \tag{33}$$

We now relate the indicator variables of eq. (24) to the above probabilities. For two lines we have the three cases distinguished by FU (1995, text and equations without number, before eq. (22))

$$\begin{aligned}
 E[\epsilon_{kl}(i)\epsilon_{k'l'}(j)] &= \delta_{i=j}p(k, i) \quad \text{if } k = k' \text{ and } l = l' \\
 E[\epsilon_{kl}(i)\epsilon_{k'l'}(j)] &= p(k, i; k, j) \quad \text{if } k = k' \\
 E[\epsilon_{kl}(i)\epsilon_{k'l'}(j)] &= p(k, i; k', j) \quad \text{else.}
 \end{aligned} \tag{34}$$

With three lines (still assuming  $k \leq k' \leq k''$ ), this extends to:

$$\begin{aligned}
 E[\epsilon_{kl}(h)\epsilon_{k'l'}(i)\epsilon_{k''l''}(j)] &= \delta_{h=i=j}p(k, i) \quad \text{if } k = k' = k'' \text{ and } l = l' = l'' \\
 E[\epsilon_{kl}(h)\epsilon_{k'l'}(i)\epsilon_{k''l''}(j)] &= \delta_{h=i}p(k, i; k'', j) \quad \text{if } k = k' \text{ and } l = l' \\
 E[\epsilon_{kl}(h)\epsilon_{k'l'}(i)\epsilon_{k''l''}(j)] &= \delta_{i=j}p(k, h; k', i) \quad \text{if } k' = k'' \text{ and } l' = l'' \\
 E[\epsilon_{kl}(h)\epsilon_{k'l'}(i)\epsilon_{k''l''}(j)] &= p(k, h; k', i; k'', j) \quad \text{else.}
 \end{aligned} \tag{35}$$

### 246 3.1.3. Averaging over line lengths

**Proposition 1.** For any  $1 \leq k, k', k'' < n$ ,  $1 \leq l \leq k$ ,  $1 \leq l' \leq k'$ ,  $1 \leq l'' \leq k''$  the following equation holds:

$$\begin{aligned}
 E[\xi_{kl}\xi_{k'l'}\xi_{k''l''}] &= \delta_{k=k'=k''}\delta_{l=l'=l''}E[\xi_{k1}] \\
 &+ \delta_{k=k'=k''}(\delta_{l=l'} + \delta_{l=l''} + \delta_{l'=l''})E[\xi_{k1}]^2 \\
 &+ \delta_{k=k'}\delta_{l=l'}E[\xi_{k1}]E[\xi_{k''1}] + \delta_{k=k''}\delta_{l=l''}E[\xi_{k1}]E[\xi_{k'1}] + \delta_{k'=k''}\delta_{l'=l''}E[\xi_{k1}]E[\xi_{k'1}] \\
 &+ (2\delta_{k=k'=k''} + \delta_{k=k'} + \delta_{k=k''} + \delta_{k'=k''} + 1)E[\xi_{k1}]E[\xi_{k'1}]E[\xi_{k''1}]
 \end{aligned} \tag{36}$$

PROOF. Let  $X$  be a random variable. It can be easily shown that, if  $X$  is exponentially distributed ( $X \sim \text{Exp}(\lambda)$ ), then the first three moments of  $X$  are  $E[X] = \frac{1}{\lambda}$ ,  $E[X^2] = \frac{2}{\lambda^2}$  and  $E[X^3] = \frac{6}{\lambda^3}$ . If  $X$  is Poisson-distributed ( $X \sim \text{Poisson}(\mu)$ ), then  $E[X] = \mu$ ,  $E[X^2] = \mu + \mu^2$  and  $E[X^3] = \mu + 3\mu^2 + \mu^3$ . In agreement with the definition of the coalescent the  $\xi_{kl}$  are distributed as  $\xi_{kl} \sim \text{Poisson}(\frac{\theta}{2}T_k)$  with  $T_k \sim \text{Exp}(\frac{2}{k(k-1)})$ .  $\xi_{kl}$  and  $\xi_{k'l'}$  are independent if

$k \neq k'$  while  $\xi_{kl}$  and  $\xi_{k'l'}$  are independent conditional on  $T_k$  for  $l \neq l'$ . We follow

here an analogous derivation as in WAKELEY (2008).

$$\begin{aligned}
 E[\xi_{kl}^3] &= E[E[\xi_{kl}^3 | T_k]] \\
 &= E\left[T_k \frac{\theta}{2} + 3\left(T_k \frac{\theta}{2}\right)^2 + \left(T_k \frac{\theta}{2}\right)^3\right] \\
 &= \frac{2}{k(k-1)} \frac{\theta}{2} + 3 \cdot 2 \frac{4}{k^2(k-1)^2} \frac{\theta^2}{4} + 6 \frac{8}{k^3(k-1)^3} \frac{\theta^3}{8} \\
 &= \frac{1}{k(k-1)} \theta + \frac{6}{k^2(k-1)^2} \theta^2 + \frac{6}{k^3(k-1)^3} \theta^3 \\
 &= E[\xi_{k1}] + 6E[\xi_{k1}]^2 + 6E[\xi_{k1}]^3
 \end{aligned}$$

$$\begin{aligned}
 E[\xi_{kl}^2 \xi_{kl'}] &= E[E[\xi_{kl}^2 \xi_{kl'} | T_k]] \\
 &= E[E[\xi_{kl}^2 | T_k] E[\xi_{kl'} | T_k]] \\
 &= E\left[\left(T_k \frac{\theta}{2} + \left(T_k \frac{\theta}{2}\right)^2\right) T_k \frac{\theta}{2}\right] \\
 &= \frac{2}{k^2(k-1)^2} \theta^2 + \frac{6}{k^3(k-1)^3} \theta^3 \\
 &= 2E[\xi_{k1}]^2 + 6E[\xi_{k1}]^3
 \end{aligned}$$

$$\begin{aligned}
 E[\xi_{kl} \xi_{kl'} \xi_{kl''}] &= E[E[\xi_{kl} \xi_{kl'} \xi_{kl''} | T_k]] \\
 &= E[E[\xi_{kl} | T_k] E[\xi_{kl'} | T_k] E[\xi_{kl''} | T_k]] \\
 &= E\left[\left(T_k \frac{\theta}{2}\right)^3\right] \\
 &= \frac{6}{k^3(1-3)^3} \theta^3 \\
 &= 6E[\xi_{k1}]^3
 \end{aligned} \tag{37}$$

$$\begin{aligned}
 E[\xi_{kl}^2 \xi_{k'l'}] &= E[\xi_{kl}^2] E[\xi_{k'l'}] \\
 &= \frac{1}{k(k-1)k'(k'-1)} \theta^2 + \frac{2}{k^2(k-1)^2 k'(k'-1)} \theta^3 \\
 &= E[\xi_{k1}] E[\xi_{k'1}] + 2E[\xi_{k1}]^2 E[\xi_{k'1}]
 \end{aligned}$$

$$\begin{aligned}
 E[\xi_{kl} \xi_{kl'} \xi_{k'l''}] &= E[\xi_{kl} \xi_{kl'}] E[\xi_{k'l''}] \\
 &= \frac{2}{k^2(k-1)^2 k'(k'-1)} \theta^3 \\
 &= 2E[\xi_{k1}]^2 E[\xi_{k'1}]
 \end{aligned}$$

$$E[\xi_{kl} \xi_{k'l'} \xi_{k''l''}] = E[\xi_{k1}] E[\xi_{k'1}] E[\xi_{k''1}]$$

247 3.1.4. Combining results

We insert now the results for averaged topologies and averaged line lengths into eq. (24):

$$\begin{aligned}
 E[\xi_h \xi_i \xi_j] &= E \left[ \left( \sum_{k=2}^n \sum_{l=1}^k \epsilon_{kl}(h) \xi_{kl} \right) \left( \sum_{k'=2}^n \sum_{l'=1}^{k'} \epsilon_{k'l'}(i) \xi_{k'l'} \right) \left( \sum_{k''=2}^n \sum_{l''=1}^{k''} \epsilon_{k''l''}(j) \xi_{k''l''} \right) \right] \\
 &= \sum_{k=2}^n \sum_{k'=2}^n \sum_{k''=2}^n \sum_{l=1}^k \sum_{l'=1}^{k'} \sum_{l''=1}^{k''} E[\epsilon_{kl}(h) \epsilon_{k'l'}(i) \epsilon_{k''l''}(j)] E[\xi_{kl} \xi_{k'l'} \xi_{k''l''}] \\
 & \hspace{25em} (38) \\
 &= \delta_{h=i=j} \sum_{k=2}^n k E[\epsilon_{k1}(h)] E[\xi_{k1}] \\
 & \quad + \sum_{k=2}^n k^2 (\delta_{h=i} E[\epsilon_{k1}(i) \epsilon_{k2}(j)] + \delta_{i=j} E[\epsilon_{k1}(j) \epsilon_{k2}(h)] + \delta_{j=h} E[\epsilon_{k1}(h) \epsilon_{k2}(i)]) E[\xi_{k1}]^2 \\
 & \quad + \sum_{k=2}^n \sum_{k'=2}^n k k' (\delta_{h=i} E[\epsilon_{k1}(i) \epsilon_{k2}(j)] + \delta_{i=j} E[\epsilon_{k1}(j) \epsilon_{k2}(h)] + \delta_{j=h} E[\epsilon_{k1}(h) \epsilon_{k2}(i)]) E[\xi_{k1}] E[\xi_{k2}] \\
 & \quad + 2 \sum_{k=2}^n k E[\epsilon_{k1}(h)] E[\xi_{k1}]^3 \\
 & \quad + \sum_{k=2}^n \sum_{k'=2}^n k k' (\delta_{h=i} E[\epsilon_{k1}(i) \epsilon_{k'1}(j)] + \delta_{i=j} E[\epsilon_{k1}(j) \epsilon_{k'1}(h)] + \delta_{j=h} E[\epsilon_{k1}(h) \epsilon_{k'1}(i)]) E[\xi_{k1}]^2 E[\xi_{k'1}] \\
 & \quad + \sum_{k=2}^n \sum_{k'=2}^n \sum_{k''=2}^n k k' k'' E[\epsilon_{k1}(h) \epsilon_{k'1}(i) \epsilon_{k''1}(j)] E[\xi_{k1}] E[\xi_{k'1}] E[\xi_{k''1}] \\
 &= \delta_{h=i=j} \sum_{k=2}^n k p(k, h) E[\xi_{k1}] \\
 & \quad + \delta_{h=i} \sum_{k=2}^n \sum_{k'=k}^n k k' (p(k, i; k', j) + p(k, j; k', i)) E[\xi_{k1}] E[\xi_{k2}] \\
 & \quad + \delta_{i=j} \sum_{k=2}^n \sum_{k'=k}^n k k' (p(k, j; k', h) + p(k, h; k', j)) E[\xi_{k1}] E[\xi_{k2}] \\
 & \quad + \delta_{j=h} \sum_{k=2}^n \sum_{k'=k}^n k k' (p(k, h; k', j) + p(k, j; k', h)) E[\xi_{k1}] E[\xi_{k2}] \\
 & \quad + \sum_{k=2}^n \sum_{k'=k}^n \sum_{k''=k'}^n k k' k'' ((p(k, h; k', i; k'', j) + p(k, i; k', j; k'', h) + p(k, j; k', h; k'', i) \\
 & \quad \quad + p(k, h; k', j; k'', i) + p(k, i; k', h; k'', j) + p(k, j; k', i; k'', h)) E[\xi_{k1}] E[\xi_{k'1}] E[\xi_{k''1}] \\
 & \hspace{25em} (39)
 \end{aligned}$$

Applying eq. (22) of (FU, 1995) to the first term of (39) yields eq. (2):

$$\sum_{k=2}^n kp(k, i)E[\xi_{k1}] = \frac{\theta}{i} = \tau_i\theta, \quad (40)$$

and applying his eq. (23) to the next three terms of (39) yields eq. (4):

$$\sum_{k=2}^n \sum_{k'=k}^n kk'(p(k, i; k', j) + p(k, j; k', i))E[\xi_{k1}]E[\xi_{k2}] = \tau_{ij}\theta^2. \quad (41)$$

We now define the remaining terms (39) as functions

$$t_x(h, i, j) = \theta^{-3} \sum_{k=2}^n \sum_{k'=k}^n \sum_{k''=k'}^n kk'k''p_x(k, h; k', i; k'', j)E[\xi_{k1}]E[\xi_{k'1}]E[\xi_{k''1}] \quad (42)$$

where  $x$  stands for  $\{aa, ab, ba^{(3)}, ba^{(2)}, ba^{(1)}, bb\}$  and finally we set

$$t_{ba}(h, i, j) = t_{ba}^{(3)}(i, j, h) + t_{ba}^{(2)}(i, h, j) + t_{ba}^{(1)}(h, i, j). \quad (43)$$

248 In the supplement we transform these functions to yield (6).

249 We offer an implementation in C++ for numerical calculation of the third  
 250 moments, given  $n$  and  $\theta$ , using the expressions (1)-(6). Just for control, we  
 251 implemented the unsimplified functions (42), too. Within rounding errors ( $<$   
 252  $10^{-12}$ ) they yield the same values as (6) for all third moments  $E[\xi_h\xi_i\xi_j]$  and  
 253 tested sample sizes  $2 \leq n \leq 17$ . With the algebraic computing software MATHE-  
 254 MATICA (WOLFRAM RESEARCH, INC., 2014) we were able to prove for the same  
 255 range of  $n$  that the expressions are exactly equivalent. The source code is con-  
 256 tained in the package “coatli”, downloadable at <http://sourceforge.net/projects/coatli>.

### 257 3.2. Proof of theorem 2

We derive the third moments of segregating sites  $S$  using the method of WATTERSON (1975). He showed (his eq. (1.3a)), that the probability generating function of  $S$  can be approximated for large population size  $N$  and small sample size  $n$  by:

$$E[s^S] = \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)}. \quad (44)$$

From general probability theory we use the formula

$$E[S^3 - 3S^2 + 2S] = \left( \frac{d^3}{ds^3} E[s^S] \right) \Big|_{s=1}. \quad (45)$$

Hence:

$$\begin{aligned} \frac{d^3}{ds^3} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} &= \frac{d^2}{ds^2} \left( \sum_{j=1}^{n-1} \frac{\frac{1}{j}\theta}{1 + \frac{1}{j}\theta(1-s)} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \right) \\ &= \frac{d}{ds} \left( \sum_{j=1}^{n-1} \frac{\frac{1}{j^2}\theta^2}{(1 + \frac{1}{j}\theta(1-s))^2} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \right. \\ &\quad \left. + \frac{d}{ds} \sum_{j=1}^{n-1} \frac{\frac{1}{j}\theta}{1 + \frac{1}{j}\theta(1-s)} \sum_{k=1}^{n-1} \frac{\frac{1}{k}\theta}{1 + \frac{1}{k}\theta(1-s)} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \right) \\ &= \sum_{j=1}^{n-1} \frac{2\frac{1}{j^3}\theta^3}{(1 + \frac{1}{j}\theta(1-s))^3} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \\ &\quad + \sum_{j=1}^{n-1} \frac{\frac{1}{j^2}\theta^2}{(1 + \frac{1}{j}\theta(1-s))^2} \sum_{k=1}^{n-1} \frac{\frac{1}{k}\theta}{1 + \frac{1}{k}\theta(1-s)} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \\ &\quad + \sum_{j=1}^{n-1} \frac{\frac{1}{j^2}\theta^2}{(1 + \frac{1}{j}\theta(1-s))^2} \sum_{k=1}^{n-1} \frac{\frac{1}{k}\theta}{1 + \frac{1}{k}\theta(1-s)} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \\ &\quad + \sum_{j=1}^{n-1} \frac{\frac{1}{j}\theta}{1 + \frac{1}{j}\theta(1-s)} \sum_{k=1}^{n-1} \frac{\frac{1}{k^2}\theta^2}{(1 + \frac{1}{k}\theta(1-s))^2} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)} \\ &\quad + \sum_{j=1}^{n-1} \frac{\frac{1}{j}\theta}{1 + \frac{1}{j}\theta(1-s)} \sum_{k=1}^{n-1} \frac{\frac{1}{k}\theta}{1 + \frac{1}{k}\theta(1-s)} \sum_{l=1}^{n-1} \frac{\frac{1}{l}\theta}{1 + \frac{1}{l}\theta(1-s)} \prod_{i=1}^{n-1} \frac{1}{1 + \frac{1}{i}\theta(1-s)}. \end{aligned} \quad (46)$$

Setting  $s = 1$  gives

$$E[S^3 - 3S^2 + 2S] = \left( 2 \sum_{i=1}^{n-1} \frac{1}{i^3} + 3 \sum_{i=1}^{n-1} \frac{1}{i^2} \sum_{i=1}^{n-1} \frac{1}{i} + \left( \sum_{i=1}^{n-1} \frac{1}{i} \right)^3 \right) \theta^3 \quad (47)$$

and inserting Watterson's results for the first and second moment (his eq. (1.4a) and (1.5a))

$$\begin{aligned} E[S] &= \sum_{i=1}^{n-1} \frac{1}{i}\theta \\ E[S^2] &= \text{Var}[S^2] + E[S]^2 = E[S] + \sum_{i=1}^{n-1} \frac{1}{i^2}\theta^2 + E[S]^2 \end{aligned}$$

yields our theorem 2:

$$\begin{aligned} E[S^3] &= E[S^3 - 3S^2 + 2S] + 3E[S^2] - 2E[S] \\ &= \left( 2 \sum_{i=1}^{n-1} \frac{1}{i^3} + 3 \sum_{i=1}^{n-1} \frac{1}{i^2} \sum_{i=1}^{n-1} \frac{1}{i} + \left( \sum_{i=1}^{n-1} \frac{1}{i} \right)^3 \right) \theta^3 + 3 \left( \sum_{i=1}^{n-1} \frac{1}{i^2} + \left( \sum_{i=1}^{n-1} \frac{1}{i} \right)^2 \right) \theta^2 + \sum_{i=1}^{n-1} \frac{1}{i} \theta. \end{aligned}$$

#### 258 4. Discussion

259 Kingman’s coalescent (KINGMAN, 1982) is an extremely useful model to de-  
260 scribe the patterns of mutations in neutral populations. For this reason, coales-  
261 cent methods were used to compute analytically the expectation and covariance  
262 of the frequency spectrum (FU, 1995). Here, we derive for the first time the  
263 third moments of the full frequency spectrum. We think, the third moments  
264 add a valuable building block to coalescent theory.

265 Beyond their fundamental interest, our results have several applications.  
266 We show how to compute analytically the bias of neutrality tests. Moreover, we  
267 describe the joint frequency spectrum for triplets of sites (fully characterising  
268 their expected haplotype structure). In turn, these results can be used to im-  
269 prove neutrality tests and approaches based on composite likelihood (KIM and  
270 STEPHAN, 2002) and Poisson random field (SAWYER and HARTL, 1992).

271 The conditional spectra can be used to characterize chromosomal inversions  
272 and introgressions (FERRETTI *et al.*, 2017). The evolution of inversions has  
273 been studied already a long time (CORBETT-DETIG and HARTL, 2012). Recent  
274 improvements of high-throughput sequencing technology allow their investiga-  
275 tion on a much larger scale (SUDMANT *et al.*, 2015). When alleles are found  
276 at intermediate frequency, it is not obvious, whether they are under balanc-  
277 ing selection, ongoing positive selection or just neutrally evolving by genetic  
278 drift (HOFFMANN and RIESEBERG, 2008). Patterns of polymorphisms in such  
279 regions may help to tackle this question. In regions with inversions, recombina-  
280 tion can be strongly inhibited (KIRKPATRICK, 2010) which allows to partition  
281 the spectrum into nested and disjoint components with respect to the inverted  
282 sequences. Nested/disjoint spectra can hence be used to extend the class of



283 frequency spectrum based tests on neutrality to cope with genomic features  
284 such as inversions and introgressions. The proper normalisation of such tests  
285 requires the knowledge of the corresponding variances and covariances, which  
286 we derived.

287 The main limit of our results is that they do not account for recombination  
288 between sites. Recombination is largely irrelevant for the spectrum of a single  
289 site, but becomes already relevant for pairs of sites. Therefore, applicability to  
290 biological data is limited to small regions or sequences with negligible recombina-  
291 tion. For this reason we present an application to the dynamics of chromosomal  
292 inversions. The patterns of mutations in these regions are naturally described in  
293 terms of the higher moments of the frequency spectrum without recombination.  
294 In particular, the expected spectrum of neutral inversions can be obtained from  
295 the second moments of the usual spectrum, as shown in FERRETTI *et al.* (2017),  
296 while the variance of the spectrum of neutral inversions requires precisely the  
297 third moments of the spectrum derived here. Applications to the detection of  
298 balancing and positive selection in chromosomal inversions and other structural  
299 variants will be presented in future publications.

300 Note that there is a close relation between the joint spectrum of multiple  
301 sites and the multi-allelic spectrum of a single locus (FERRETTI *et al.*, 2017). In  
302 fact, at low mutation rates, we can consider the multiple sites as a single locus  
303 with multiple alleles, and retrieve the multi-allelic spectrum for the locus by  
304 considering the frequencies of the  $m + 1$  alleles that result from the  $m$  polymor-  
305 phic sites. In this light, our results can be used to derive the full quadri-allelic  
306 frequency spectrum. This could be applied to several multiallelic variants, the  
307 more relevant being nucleotide polymorphism (which have at most four alle-  
308 les A,C,G,T). Related results can be found in JENKINS and SONG (2011) and  
309 BHASKAR *et al.* (2012).

310 The results in this paper apply to a sample of size  $n$  much smaller than the  
311 size of the population. The spectrum for large samples converges to the contin-  
312 uous population spectrum for triplets of sites. It would be interesting to derive  
313 analytically simple expressions for such a spectrum, similarly to FERRETTI *et al.*

314 (2017). However, our expressions contain many explicit sums that prevent a di-  
315 rect computation. A further simplification of the expressions provided in this  
316 paper would be helpful.

### 317 **Acknowledgment**

318 We thank Iulia Dahmer and Götz Kersting for insightful discussions. AK  
319 was supported by a grant of the German Science Foundation (DFG-SFB680) to  
320 T. Wiehe (University of Cologne).

321 ACHAZ, G., 2009 Frequency Spectrum Neutrality Tests: One for All and All  
322 for One. *Genetics* **183**: 249–258.

323 BHASKAR, A., J. A. KAMM, and Y. S. SONG, 2012 Approximate Sampling  
324 Formulas for General Finite-Alleles Models of Mutation. *Advanced Applied*  
325 *probability* **44**: 408–428.

326 CORBETT-DETIG, R. B., and D. L. HARTL, 2012 Population Genomics of  
327 Inversion Polymorphisms in *Drosophila melanogaster*. *PLoS Genetics* **8**.

328 DAHMER, I., and G. KERSTING, 2015 The internal branch lengths of the king-  
329 man coalescent. *Annals of Applied Probability* **25**: 1325–1348.

330 DURRETT, R., 2008 *Probability Models for DNA Sequence Evolution*. Springer,  
331 2nd edition.

332 EWENS, W., 1979 *Mathematical Population Genetics*. Springer Verlag.

333 FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection.  
334 *Genetics* **155**: 1405–13.

335 FERRETTI, L., A. KLASSMANN, T. WIEHE, S. RAMOS-ONZINS, and G. AC-  
336 HAZ, 2017 The expected neutral frequency spectrum of two linked sites.  
337 arXiv:1604.06713 .

338 FERRETTI, L., M. PEREZ-ENCISO, and S. E. RAMOS-ONSINS, 2010 Optimal  
339 neutrality tests based on the frequency spectrum. *Genetics* **186**: 353–65.

- 340 FISHER, R. A., 1930 The Distribution of Gene Ratios for Rare Mutations.  
341 Proceedings of the Royal Society of Edinburgh : 205,220.
- 342 FU, Y.-X., 1995 Statistical Properties of Segregating Sites. Theoretical Popu-  
343 lation Biology **48**: 172–197.
- 344 FU, Y.-X., and W.-H. LI, 1993 Maximum likelihood estimation of population  
345 parameters. Genetics **134**: 1261–70.
- 346 GRIFFITHS, R., and S. TAVARÉ, 2003 The genealogy of a neutral mutation. In  
347 *Highly Structured Stochastic Systems*. Oxford university press, 393–412.
- 348 HEIN, J., M. SCHIERUP, and C. WIUF, 2004 *Gene Genealogies, Variation and*  
349 *Evolution: A primer in coalescent theory*. Oxford university press.
- 350 HOFFMANN, A. A., and L. H. RIESEBERG, 2008 Revisiting the Impact of Inver-  
351 sions in Evolution: From Population Genetic Markers to Drivers of Adaptive  
352 Shifts and Speciation? Annual review of ecology, evolution, and systematics  
353 **39**: 21–42.
- 354 HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic re-  
355 combination. Theoretical population biology **23**: 183–201.
- 356 HUDSON, R. R., 1991 Gene genealogies and the coalescent process. In *Oxford*  
357 *Surveys in Evolutionary Biology*. Oxford university press, 1–44.
- 358 HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model  
359 of genetic variation. Bioinformatics **18**: 337–338.
- 360 HUDSON, R. R., 2015 A New Proof of the Expected Frequency Spectrum under  
361 the Standard Neutral Model. Plos One **10**: e0118087.
- 362 JANSON, S., and G. KERSTING, 2011 On the total external length of the evolu-  
363 ting Kingman coalescent. Electronic Journal of Probability **80**: 2203–2218.
- 364 JENKINS, P. A., and Y. S. SONG, 2011 The effect of recurrent mutation on the  
365 frequency spectrum of a segregating site and the age of an allele. Theoretical  
366 Population Biology **80**: 158–173.

- 367 KIM, Y., and W. STEPHAN, 2002 Detecting a local signature of genetic hitch-  
368 hiking along a recombining chromosome. *Genetics* **160**: 765–777.
- 369 KIMURA, 1964 Diffusion Models in Population Genetics. *Journal of Applied*  
370 *Probability* **1**: 177–232.
- 371 KINGMAN, J. F. C., 1982 The coalescent. *Stochastic processes and their appli-*  
372 *cations* **13**: 235–248.
- 373 KIRKPATRICK, M., 2010 How and Why Chromosome Inversions Evolve. *PLoS*  
374 *Biology* **8**: e1000501.
- 375 NIELSEN, R., C. D. BUSTAMANTE, A. G. CLARK, S. GLANOWSKI, T. B.  
376 SACKTON, *et al.*, 2005 A scan for positively selected genes in the genomes of  
377 humans and chimpanzees. *PLoS biology* **3**: e170.
- 378 SARGSYAN, O., 2015 An analytical framework in the general coalescent tree  
379 setting for analyzing polymorphisms created by two mutations. *Journal of*  
380 *Mathematical Biology* **70**: 913–956.
- 381 SAWYER, S., and D. HARTL, 1992 Population genetics of polymorphism and  
382 divergence. *Genetics* .
- 383 SUDMANT, P. H., T. RAUSCH, E. J. GARDNER, R. E. HANDSAKER, A. ABY-  
384 ZOV, *et al.*, 2015 An integrated map of structural variation in 2,504 human  
385 genomes. *Nature* **526**: 75–81.
- 386 TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis  
387 by DNA polymorphism. *Genetics* **123**: 585–95.
- 388 TAVARÉ, S., 1984 Line-of-descent and genealogical processes, and their appli-  
389 cations in population genetics models. *Theoretical population biology* **26**:  
390 119–64.
- 391 VAN ERP, N., and P. VAN GELDER, 2007 On the moments of functions of ran-  
392 dom variables using multivariate taylor expansion, part i. In *5th International*  
393 *Probabilistic Workshop-Taerwe & Proske (eds), Ghent.*

- 394 WAKELEY, J., 2008 *Coalescent theory: an introduction*. W. H. Freeman.
- 395 WATTERSON, G., 1975 On the number of Segregating Sites in Genetical Models  
396 without Recombination. *Theoretical Population Biology* **7**: 256–276.
- 397 WOLFRAM RESEARCH, INC., 2014 *Mathematica 10.0*.
- 398 ZENG, K., Y.-X. FU, S. SHI, and C.-I. WU, 2006 Statistical tests for detecting  
399 positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431–9.

400 **List of Figures**

401 1 Skewness of neutrality tests for sample size  $n = 50$  (top) and  
402  $n = 500$  (bottom). The analytical skewness was obtained by  
403 eq. (19). For simulations, the skewness was estimated by  $\hat{\gamma} =$   
404  $\frac{\frac{1}{n} \sum_i (x_i - \bar{x})^3}{(\frac{1}{n} \sum_i (x_i - \bar{x})^2)^{\frac{3}{2}}}$  over  $10^6$  genealogies. The test values were calcu-  
405 lated using the true  $\theta$  (green points) and Wattersons estimator  
406  $\hat{\theta} = \frac{S}{a_n}$  (blue points), respectively. . . . . 31

407 2 The bias of the tests mentioned in table 1 with sample size  $n = 50$   
408 (top) and  $n = 500$  (bottom). Shown are the values of our analyt-  
409 ical approximation and numerical data, obtained by simulation  
410 with 'ms', averaged over  $10^6$  genealogies. . . . . 32

411 3 The relative error between expected (analytical) and observed  
412 (estimated from simulated data) third moments. We computed  
413 errors  $e = \max_{h,i,j} \frac{|E_{hij} - O_{hij}|}{E_{hij}}$  where  $O_{hij}$  was averaged over  $10^3$  til  
414  $10^9$  genealogies (number of samples). The figure shows the average  
415 over 100 of these errors. The colours indicate different sample  
416 sizes  $n$ . . . . . 33

417 4 The expected values of all third moments for  $n = 5$ ,  $\theta = 1$  and the  
418 respective contributions of the linear, quadratic and cubic terms. 34

419 5 Comparison between unconditional and conditional covariances.  
420 Left: unconditional covariances  $Cov[\xi_i, \xi_j]$  for sample size  $n = 10$ .  
421 The remainder graphs show the covariances between mutations  
422 conditional on a mutation of size  $k = 10$  in a sample of size  
423  $n = 20$ . The left middle shows the covariances between muta-  
424 tions nested within the focal mutation, the right middle the  
425 covariances of mutations both disjoint and the rightmost the co-  
426 variance between nested and disjoint mutations. . . . . 35

427 6 From (DAHMER and KERSTING, 2015) follows, that mutations of  
428 small size within a large sample are approximately independent.  
429 Shown is the ratio of our exact results ( $\tau_{hij}\theta^3$ ) to the asymptotic  
430 approximation (eq. (23)) for small fixed indices  $1 \leq h, i, j \leq 4$   
431 and varying sample size  $n$ . Left: all indices are the same; middle:  
432 two indices differ; right: all indices differ. . . . . 36

433 7 How do we calculate the expected number of mutations of size  
434 2? For a tree with this topology  $\mathcal{T}$  only the lines  $\xi_{21}$ ,  $\xi_{22}$  and  
435  $\xi_{33}$  have two descendants in the sample. Thus we have  $\epsilon_{21}(2) =$   
436  $\epsilon_{22}(2) = \epsilon_{33}(2) = 1$  and all other  $\epsilon_{kl}(2)$  are zero. It follows, that  
437  $E[\xi_2 | Topology = \mathcal{T}] = E[\xi_{21}] + E[\xi_{22}] + E[\xi_{33}]$ .  $E[\xi_{kl}]$  is the expected  
438 amount of mutations on the line  $\xi_{kl}$  which is proportional to its  
439 length. Averaging over all topologies yields  $E[\xi_2]$ . . . . . 37

440 8 The possible hierachical relationships between three lines of a  
441 coalescent tree and their corresponding probabilities. . . . . 38

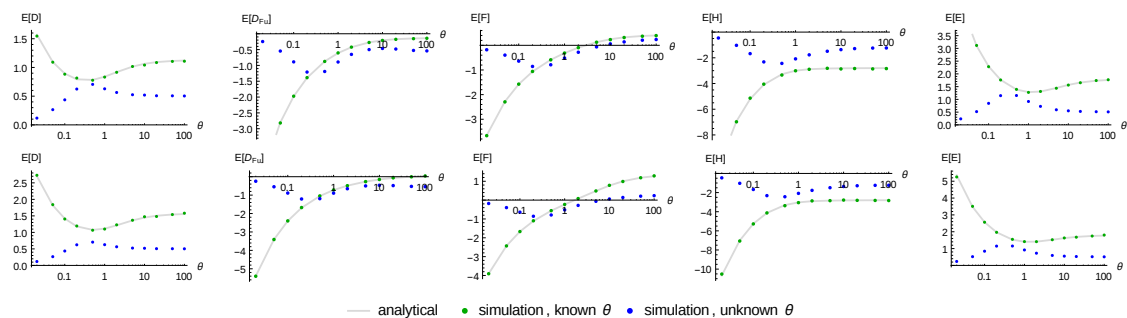


Figure 1: Skewness of neutrality tests for sample size  $n = 50$  (top) and  $n = 500$  (bottom). The analytical skewness was obtained by eq. (19). For simulations, the skewness was estimated by  $\hat{\gamma} = \frac{\frac{1}{n} \sum_i (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_i (x_i - \bar{x})^2\right)^{\frac{3}{2}}}$  over  $10^6$  genealogies. The test values were calculated using the true  $\theta$  (green points) and Wattersons estimator  $\hat{\theta} = \frac{S}{a_n}$  (blue points), respectively.

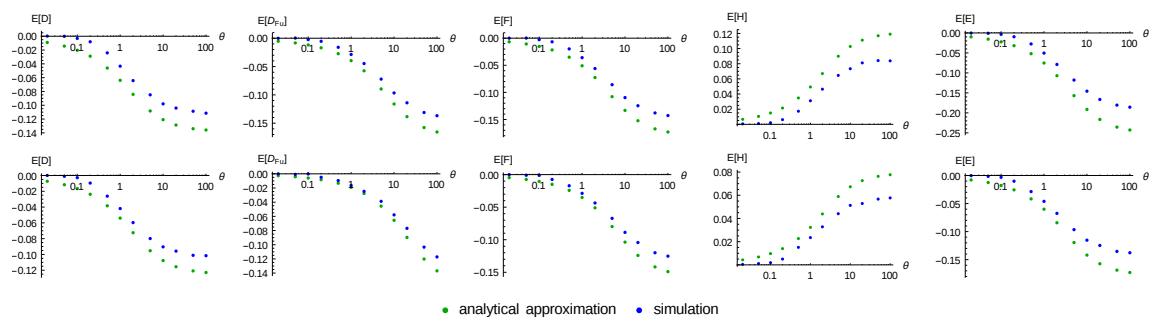


Figure 2: The bias of the tests mentioned in table 1 with sample size  $n = 50$  (top) and  $n = 500$  (bottom). Shown are the values of our analytical approximation and numerical data, obtained by simulation with 'ms', averaged over  $10^6$  genealogies.



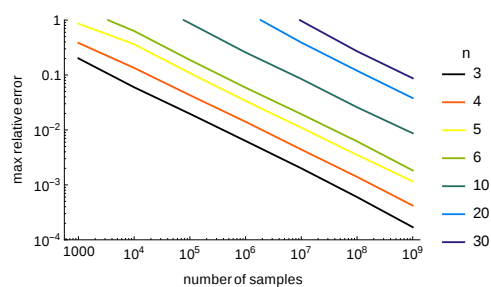


Figure 3: The relative error between expected (analytical) and observed (estimated from simulated data) third moments. We computed errors  $e = \max_{h,i,j} \frac{|E_{hij} - O_{hij}|}{E_{hij}}$  where  $O_{hij}$  was averaged over  $10^3$  til  $10^9$  genealogies (number of samples). The figure shows the average over 100 of these errors. The colours indicate different sample sizes  $n$ .

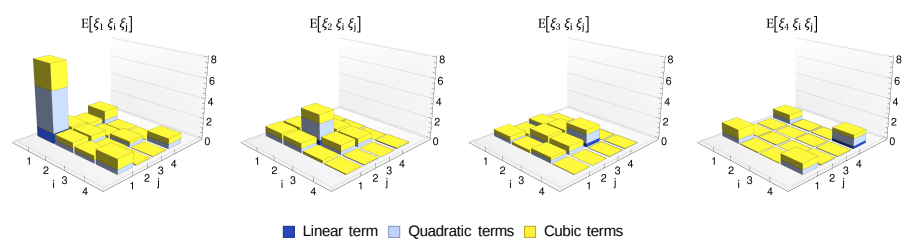


Figure 4: The expected values of all third moments for  $n = 5$ ,  $\theta = 1$  and the respective contributions of the linear, quadratic and cubic terms.

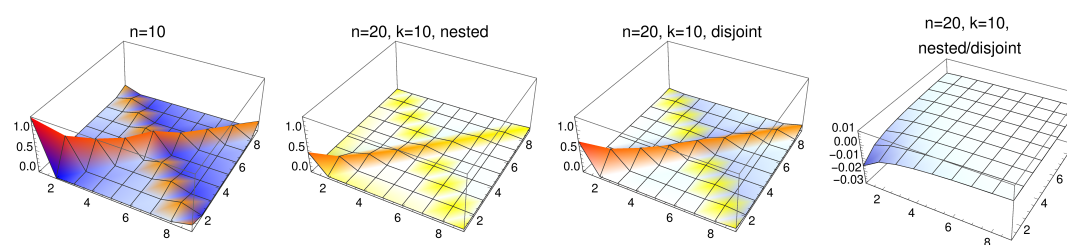


Figure 5: Comparison between unconditional and conditional covariances. Left: unconditional covariances  $Cov[\xi_i, \xi_j]$  for sample size  $n = 10$ . The remainder graphs show the covariances between mutations conditional on a mutation of size  $k = 10$  in a sample of size  $n = 20$ . The left middle shows the covariances between mutations nested within the focal mutation, the right middle the covariances of mutations both disjoint and the rightmost the covariance between nested and disjoint mutations.

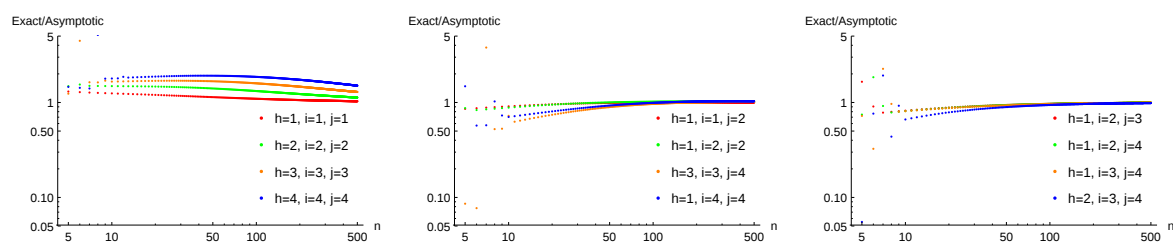


Figure 6: From (DAHMER and KERSTING, 2015) follows, that mutations of small size within a large sample are approximately independent. Shown is the ratio of our exact results ( $\tau_{hij}\theta^3$ ) to the asymptotic approximation (eq. (23)) for small fixed indices  $1 \leq h, i, j \leq 4$  and varying sample size  $n$ . Left: all indices are the same; middle: two indices differ; right: all indices differ.

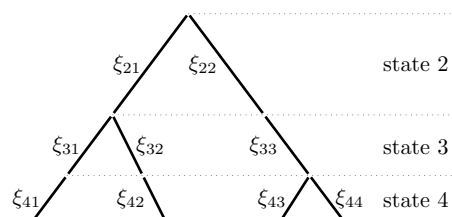


Figure 7: How do we calculate the expected number of mutations of size 2? For a tree with this topology  $\mathcal{T}$  only the lines  $\xi_{21}$ ,  $\xi_{22}$  and  $\xi_{33}$  have two descendants in the sample. Thus we have  $\epsilon_{21}(2) = \epsilon_{22}(2) = \epsilon_{33}(2) = 1$  and all other  $\epsilon_{kl}(2)$  are zero. It follows, that  $E[\xi_2 | \text{Topology} = \mathcal{T}] = E[\xi_{21}] + E[\xi_{22}] + E[\xi_{33}]$ .  $E[\xi_{kl}]$  is the expected amount of mutations on the line  $\xi_{kl}$  which is proportional to its length. Averaging over all topologies yields  $E[\xi_2]$ .

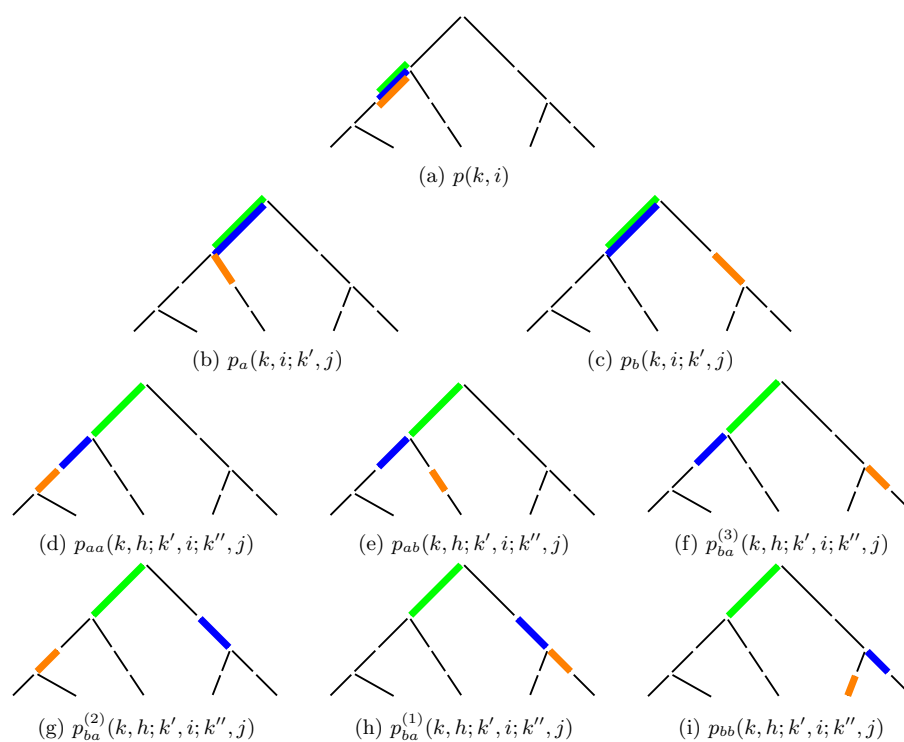


Figure 8: The possible hierarchical relationships between three lines of a coalescent tree and their corresponding probabilities.

442 **List of Tables**

443     1     Weights and references of the analysed neutrality tests. . . . . 40

Test	weights $\Omega_i$	reference
$D_{(Tajima)}$	$(n-i)/\binom{n}{2} - 1/ia_n$	TAJIMA (1989)
$D_{(Fu\&Li)}$	$1/ia_n - \delta_{i,1}$	FU and LI (1993)
$F_{(Fu\&Li)}$	$(n-i) - \delta_{i,1}$	FU and LI (1993)
$H_{(Fay\&Wu)}$	$(n-2i)/\binom{n}{2}$	FAY and WU (2000)
$E_{(Zeng)}$	$1/(n-1) - 1/ia_n$	ZENG <i>et al.</i> (2006)

Table 1: Weights and references of the analysed neutrality tests.