The Site Frequency Spectrum under Finite and Time-Varying Mutation Rates

Andrew Melfi and Divakar Viswanath

Department of Mathematics, University of Michigan (melfi/divakar@umich.edu).

Abstract

The diversity in genomes is due to the accumulation of mutations and the site frequency spectrum (SFS) is a popular statistic for summarizing genomic data. The current coalescent algorithm for calculating the SFS for a given demography assumes the $\mu \rightarrow 0$ limit, where μ is the mutation probability (or rate) per base pair per generation. The algorithm is applicable when μN , N being the haploid population size, is negligible. We derive a coalescent based algorithm for calculating the SFS that allows the mutation rate $\mu(t)$ as well as the population size $N(t)$ to vary arbitrarily as a function of time. That algorithm shows that the probability of two mutations in the genealogy becomes noticeable already for $\mu = 10^{-8}$ for samples of $n = 10^5$ haploid human genomes and increases rapidly with μ . Our algorithm calculates the SFS under the assumption of a single mutation in the genealogy, and the part of the SFS due to a single mutation depends only mildly on the finiteness of μ . However, the dependence of the SFS on variation in μ can be substantial for even $n = 100$ samples. In addition, increasing and decreasing mutation rates alter the SFS in different ways and to different extents.

Introduction

Haldane [1947] noted that the gene for hemophilia appeared to mutate more rapidly in men than in women (also see [Rahbari et al., 2016]). Since then, a great deal of evidence has accumulated that the mutation rate varies across the human genome. CpG junctions are well-known to have particularly high mutation rates [Ehrlich and Wang, 1981]. Some of the variation in the mutation rate across the human genome appears to be related to correlations between single nucleotide polymorphisms found in humans as well as other nearby primates [Harpak et al., 2016, Hodgkinson et al., 2009, Hodgkinson and Eyre-Walker, 2011, Johnson and Hellmann, 2011]. A variety of statistical models of the variation of the mutation rate have been proposed and examined [Aggarwala and Voight, 2016, Carlson et al., 2018, Eyre-Walker and Eyre-Walker, 2014, Hodgkinson and Eyre-Walker, 2011, Michaelson et al., 2012]. The mutation rate can vary even between human populations [Harris, 2015, Mathieson and Reich, 2017, Narasimhan et al., 2017].

The site frequency spectrum (SFS) is commonly used to summarize the effect of mutations across the genome. For a haploid sample of size n , the SFS consists of the probability that j of the samples carry the mutant allele, for $j = 1, \dots, n-1$, at a polymorphic site. If the mutation rate itself varies widely across the genome, it is essential to know how the mutation rate affects the SFS. In this article, we derive an algorithm to calculate the SFS with mutation rates as

well as population sizes allowed to vary in an arbitrary manner. The mutation rate is assumed to be $\mu(t)$ per base pair per generation at time t and the haploid population size is assumed to be $N(t)$. The algorithm relies on the coalescent approximation to genealogies [Durrett, 2008]. In particular, n samples are assumed to experience a binary merger according to a Poisson process of rate $n(n-1)/2N(t)$. The samples are hit with mutations by an independent Poisson process of rate $\mu(t)n$.

An algorithm for calculating the SFS assuming μN to be negligible and μ to be constant in time is due to Polanski and Kimmel [2003]. The Polanski-Kimmel algorithm, which relies on the earlier work of Griffiths and Tavaré [1998] as well as Polanski et al. [2003], is based on the internal structure of the coalescent genealogy. In particular, the algorithm relies on the expected branch length of the genealogy with exactly b descendants for $b = 1, \dots, n-1$.

Our algorithm allows $\mu(t)$ to be finite and varying in time and is also based on the coalescent approximation. However, it pays no attention to the internal structure of the genealogy. The algorithm is more Markovian in spirit and is partly based on the ideas in our earlier analytic work [Melfi and Viswanath, 2018a,b].

Harpak et al. [2016] have presented data analysis showing that samples of size $n \approx 10^5$ have experienced more than one mutation at several polymorphic sites with $\mu \in [10^{-9}, 10^{-7}]$. Our algorithm can calculate the probability that a polymorphic site has experienced more than one mutation exactly. Using the demography inferred by Harpak et al. [2016], we precisely delineate the probability of more than one mutation in the genealogy.

When n and μ are large enough that a polymorphic site has been hit with either one, two, or more mutations, the SFS is a mixture of the SFS due to a single mutation and the SFS due to two or more mutations. Our calculations imply that the effects described by Harpak et al. [2016], such as the change in the profile of rare alleles at sites of higher μ , are mostly due to two or more mutations in the genealogy, which is in agreement with their conclusions.

Beginning with the work of Hwang and Green [2004], a number of authors have questioned the constancy of $\mu(t)$ with respect to t [Kim et al., 2006, Moorjani et al., 2016b,a, Narasimhan et al., 2017, Scally and Durbin, 2012]. The germ line mutation rate is known to depend on the number of cell divisions experienced by the germ line [Gao et al., 2016, Amster and Sella, 2016]. The number of cell divisions in the germ line is greater in the human male than the human female, and in the male it increases with age. The dependence on the number of cell divisions could be due to errors during either genome replication or DNA repair [Gao et al., 2016], and some of the de novo mutations are shared between siblings in a mosaic pattern [Rahbari et al., 2016]. We use our algorithm to illustrate how increasing and decreasing mutation rates alter the SFS.

The coalescent and the diffusion equation often provide alternative routes to the same results. Accordingly, the SFS can be computed using the diffusion equation [Griffiths, 2003, Gutenkunst et al., 2009, Sawyer and Hartl, 1992]. For the possibility of handling varying mutation rates using the diffusion equation, see [Steinrücken et al., 2016, Živković et al., 2015]. In the diffusion approach, the transition probabilities are first obtained and the SFS is computed using the transition probabilities. The sample size n enters only during the latter step. Therefore, it may appear as if the diffusion equation can calculate SFS for even large n with not much more trouble than for small n . However, for large n , the transition probabilities have to be calculated more accurately and with greater resolution.

Calculating the SFS under Finite and Varying Mutation Rates

Suppose the population size $N(t) \equiv N$ and mutation rate $\mu(t) \equiv \mu$ are both constant. If a site is polymorphic, the probability that j out of n samples carry the mutant allele converges to

$$\frac{1/j}{\mathcal{H}_{n-1}}, \quad (1)$$

where $\mathcal{H}_{n-1} = 1 + 1/2 + \dots + 1/(n-1)$, in the limit $\mu N \rightarrow 0$ [Durrett, 2008]. Before presenting the full algorithm, we treat the case where N and μ are both constant but without the assumption of μN being negligible. The extension of the classical result (1) that we derive brings out some of the main ideas in a relatively transparent way.

If the coalescent genealogy is sectioned at some fixed time in the past, we will refer to the lineages present at that time in the past as ancestral samples, following our earlier usage [Melfi and Viswanath, 2018a,b]. Suppose the number of ancestral samples is k . Because the Poisson process of rate $k(k-1)/2N$ that produces a binary merger in the ancestral sample and the Poisson process of rate μk that hits the sample with a mutation are independent, the probability that the next event in the genealogy is a binary merger is

$$\frac{k(k-1)/2N}{k(k-1)/2N + 2k\mu} = \frac{k-1}{k-1 + 2N\mu}. \quad (2)$$

Correspondingly, the probability that the next event is a mutation is

$$\frac{2N\mu}{k-1 + 2N\mu}. \quad (3)$$

It follows that the probability $q_0(k)$ that k samples coalesce without being hit by a mutation is

$$q_0(k) = \prod_{j=2}^k \frac{j-1}{j-1 + 2N\mu}.$$

In more detail, for ancestral sample sizes $j = 2, \dots, k$, a binary merger must precede a mutation, which occurs with a probability given by (2) (with $k \leftarrow j$) for each $j = 2, \dots, k$.

Similarly, the probability $q_1(k)$ that a sample of size k coalesces after experiencing exactly one mutation is

$$q_1(k) = \prod_{j=2}^k \frac{j-1}{j-1 + 2N\mu} \times \sum_{j=2}^k \frac{2N\mu}{j-1 + 2N\mu}.$$

In more detail, the probability that a sample of size k coalesces after experiencing exactly one mutation when the ancestral sample size is j but with no other mutations in the genealogy is

$$\prod_{\substack{\ell=2 \\ \ell \neq j}}^k (\ell-1)/(\ell-1 + 2N\mu) \times \frac{2N\mu}{j-1 + 2N\mu} \times \frac{j-1}{j-1 + 2N\mu}.$$

The first factor occurs because the first event to hit an ancestral sample of size $\ell \neq j$ is a binary merger. The second factor occurs because the first event experienced by an ancestral sample of size j must be a mutation (whose probability is given by (3) with $k \leftarrow j$) and the

third factor because the sample of size j then experiences a binary merger (whose probability is given by (2) with $k \leftarrow j$). The formula for $q_1(k)$ is obtained by summing over $j = 2, \dots, k$.

The condition or event that n samples coalesce with exactly one mutation in the genealogy will be denoted by $\mathcal{C}_n$. The definition of $\mathcal{C}_n$ will be changed slightly after we introduce the concept of an ancestral lens. Conditioned on $\mathcal{C}_n$, the probability that a mutation event occurs in the genealogy when the sample size is k is given by

$$\frac{\frac{1}{k-1+2N\mu}}{\sum_{j=2}^n \frac{1}{j-1+2N\mu}}.$$

Using the coalescent propagators derived earlier [Griffiths and Tavaré, 1998, Melfi and Viswanath, 2018b], we may deduce the probability of j mutants in the sample under the condition $\mathcal{C}_n$ to be

$$\frac{1}{\sum_{k=2}^n \frac{1}{k-1+2N\mu}} \sum_{k=2}^n \frac{1}{k-1+2N\mu} \times \frac{(n-k)^{j-1}(k-1)}{(n-1)^j}, \quad (4)$$

where $\mathbf{a}^{\mathbf{b}}$ denotes the falling power $\mathbf{a}(\mathbf{a}-1)\dots(\mathbf{a}-\mathbf{b}+1)$. The SFS (1) is obtained by substituting $\mu = 0$ in this formula.

There is no obvious way to evaluate this formula for $j = 1, \dots, n-1$ in $\mathcal{O}(n)$ arithmetic operations for $\mu > 0$, although the $\mu = 0$ case given by (1) can be evaluated in $\mathcal{O}(n)$ arithmetic operations. The formula (4) for the SFS under the condition $\mathcal{C}_n$ can be cast in the form of a recurrence and the SFS evaluated using $\mathcal{O}(n^2)$ operations [Melfi and Viswanath, 2018a].

This derivation of the SFS conditioned on $\mathcal{C}_n$ for constant N and μ relies on $q_0(k)$, the probability that k samples coalesce with zero mutations, and $q_1(k)$, the probability that k samples coalesce with exactly one mutation in their genealogy. The algorithm with varying $N(t)$ and $\mu(t)$ uses a similar approach. The more general algorithm relies on the concept of an ancestral lens, to which we now turn, to reduce the number of arithmetic operations.

Ancestral and calendar lenses

We denote ancestral time by τ (generations), with the current epoch being $\tau = 0$. If the number of samples is n , let $p(k, \tau)$ be the probability that the number of ancestral samples at time τ is k . The ancestral lens is defined as the set of all (k, τ) such that $p(k, \tau) \geq \epsilon_{lens}$. The tolerance ϵ_{lens} is taken to be 10^{-40} . All sample sizes outside the ancestral lens have a negligible probability and may be ignored without affecting any calculations for the current sample at $\tau = 0$. Significant savings may be realized by confining calculations to the ancestral lens.

The probability that the ancestral sample is of size k at time τ and undergoes a binary merger in the interval $[\tau, \tau + d\tau]$ is

$$p(k, \tau) \frac{k(k-1)}{2N(\tau)} d\tau + \mathcal{O}(d\tau^2).$$

Likewise, the probability that the ancestral sample is of size $k+1$ and undergoes a binary merger in $[\tau, \tau + d\tau]$ is

$$p(k+1, \tau) \frac{(k+1)k}{2N(\tau)} d\tau + \mathcal{O}(d\tau^2).$$

Fig. 1: The ancestral lens with $n = 10^5$ and demographic model 2. The second plot shows the probability of (non)coalescence as a function of ancestral time τ .

Therefore,

$$p(k, \tau + d\tau) - p(k, \tau) = -p(k, \tau) \frac{k(k-1)}{2N(\tau)} d\tau + p(k+1, \tau) \frac{(k+1)k}{2N(\tau)} d\tau + \mathcal{O}(d\tau^2).$$

In the limit $d\tau \rightarrow 0$, we obtain the differential equation

$$\frac{dp(k, \tau)}{d\tau} = -\frac{k(k-1)}{2N(\tau)} p(k, \tau) + \frac{k(k+1)}{2N(\tau)} p(k+1, \tau). \quad (5)$$

To calculate the ancestral lens, this differential equation is initialized with $p(n, 0) = 1$ and $p(k, 0) = 0$ for $k \neq n$. The numerical method used for computing $k_{min}(\tau)$ and $k_{max}(\tau)$ such that $p(k, \tau) < \epsilon_{lens}$ for $k \notin [k_{min}(\tau), k_{max}(\tau)]$ is described in the appendix. The functions $k_{min}(\tau)$ and $k_{max}(\tau)$ are the boundaries of the ancestral lens.

Figure 1 shows the ancestral lens for a demographic model. The considerable savings realized by confining calculations to the ancestral lens are obvious from that figure. We work with three demographic models:

- Demographic model 0: N is assumed to be constant and equal to 2×10^4 .
- Demographic model 1: $N(\tau) = N_0 e^{-r\tau}$, with $N_0 = 8 \times 10^6$ and $r = 0.017$, for $\tau \in [0, 367.8]$ and $N(\tau) = N_0 e^{-r \cdot 367.8} \approx 15403$ for $\tau \geq 367.8$. This model is from [Nelson et al., 2012].
- Demographic model 2: $N(\tau) = N_0 e^{-r\tau}$, with $N_0 = 8 \times 10^6$ and $r = 0.0538$, for $\tau \in [0, 119.47]$ and $N(\tau) = N_0 e^{-r \cdot 119.47} \approx 12932$ for $\tau \geq 119.47$. This model is from [Harpak et al., 2016].

Figure 1 uses model 2. In that figure, the stopping time $\tau = T$ for the ancestral lens is determined using the criterion $1 - p(1, T) \geq 10^{-3}$, which ensures that the entire sample has coalesced with probability greater than 99.9%. Such a criterion makes T very large, however, as shown in the figure. It is better to stop the lens when μ and N have become constant. The entire section of the lens at the stopping time T can be initialized using exact formulas for q_0 , q_1 , and the SFS.

The computation of the ancestral lens is forward in ancestral time τ . However, the computation of q_0 , q_1 , and the SFS march forward in calendar time. Therefore, the ancestral lens

Fig. 2: The ancestral and calendar lenses for $n = 1000$ samples under demographic model 1.

must be flipped to a calendar time lens as shown in Figure 2. In both Figures 1 and 2, k^* is the most likely size of the ancestral sample. We take the stopping time $\tau = T$ to be the origin $t = 0$ of calendar time so that the current epoch is $t = T$.

Computation of $q_0(k, t)$ and $q_1(k, t)$

The probability that k samples at calendar time t , $t \in [0, T]$, coalesce within the lens or exit the lens without being hit by a mutation is denoted $q_0(k, t)$. The probability of exiting the lens is very low and we regard any exit from the lens as equivalent to coalescence. Any loss of accuracy will be below the level of machine precision. Thus, $q_0(k, t) = 1$ for $k \notin [k_{\min}(t), k_{\max}(t)]$. In addition, $q_0(1, t) = 1$.

Suppose there are k ancestral samples at time t . If we go back to time $t - dt$ and ignore dt^2 terms, we have the following possibilities:

- The sample is hit with a mutation with probability $\mu(t)k dt$.
- There is a binary merger with probability $k(k - 1) dt / 2N(t)$.
- There is neither a mutation nor a binary merger with probability $1 - \mu(t)k dt - k(k - 1) dt / 2N(t)$.

Because we are ignoring dt^2 terms, these three possibilities are disjoint and exhaustive.

We have

$$q_0(k, t) = \left(1 - \mu(t)k dt - \frac{k(k - 1) dt}{2N(t)}\right) q_0(k, t - dt) + \frac{k(k - 1)}{2N(t)} q_0(k - 1, t - dt).$$

If the k samples are neither hit by a mutation nor experience a binary merger in $[t - dt, t]$, we require that k samples at time $t - dt$ coalesce (or exit the lens) without a mutation. If the samples experience a binary merger, we require that $k - 1$ samples at time $t - dt$ coalesce (or exit the lens) without a mutation. In the $dt \rightarrow 0$ limit, we have the following differential equation:

$$\frac{dq_0(k, t)}{dt} = - \left(\mu(t)k + \frac{k(k - 1)}{2N(t)} \right) q_0(k, t) + \frac{k(k - 1)}{2N(t)} q_0(k - 1, t). \quad (6)$$

The differential equation is solved for $t \in [0, T]$ as an initial value problem. For $k(0) \leq k \leq k_{max}(0)$, $q_0(k, 0)$ is initialized using the exact formula for $q_0(k)$ when μ and N are constant. After solving the initial value problem, we obtain $q_0(n, T)$, which is the probability that n samples at the current epoch coalesce (or exit the lens) without being hit by a mutation.

Similarly, let $q_1(k, t)$ be the probability that k ancestral samples at time t coalesce (or exit the lens) while being hit by exactly one mutation. In this case, the same three disjoint and exhaustive possibilities imply

$$q_1(k, t) = \left(1 - \mu(t)k dt - \frac{k(k-1)dt}{2N(t)}\right) q_1(k, t-dt) + \mu(t)k dt q_0(k, t-dt) + \frac{k(k-1)}{2N(t)} q_1(k-1, t).$$

If there is neither a binary merger nor a mutation in $[t-dt, t]$, we require that k samples at time $t-dt$ coalesce (or exit the lens) with exactly one mutation. If there is a mutation event, we require that k samples at time $t-dt$ coalesce (or exit the lens) without suffering a mutation. Finally, if there is binary merger, we require that $k-1$ samples at time t coalesce (or exit the lens) while suffering exactly one mutation.

In the limit $dt \rightarrow 0$, we have the differential equation

$$\frac{dq_1(k, t)}{dt} = -\left(\mu(t)k + \frac{k(k-1)}{2N(t)}\right) q_1(k, t) + \mu(t)k q_0(k, t) + \frac{k(k-1)}{2N(t)} q_1(k-1, t). \quad (7)$$

This differential equation too is solved as an initial value problem from $t = 0$ to $t = T$. For $k \in [k_{min}(0), k_{max}(0)]$, $q_1(k, 0)$ is initialized using the exact formula for $q_1(k)$ with constant μ and N . If $k \notin [k_{min}(t), k_{max}(t)]$, then $q_1(k, t) = 0$. In addition, $q_1(1, t) = 0$ for $t \in [0, T]$.

If $q_2(k, t)$ is the probability that k ancestral samples at time t coalesce with exactly 2 mutations, $q_2(k, t)$ satisfies an analogous differential equation:

$$\frac{dq_2(k, t)}{dt} = -\left(\mu(t)k + \frac{k(k-1)}{2N(t)}\right) q_2(k, t) + \mu(t)k q_1(k, t) + \frac{k(k-1)}{2N(t)} q_2(k-1, t).$$

If N and μ are constant

$$q_2(k) = q_0(k) \sum_{2 \leq i \leq j \leq k} \frac{(2N\mu)^2}{(i-1+2N\mu)(j-1+2N\mu)}.$$

The initial value problem for $q_2(k, t)$ may be initialized using $q_2(k, 0) = q_2(k)$ with $N = N(0)$ and $\mu = \mu(0)$.

Computation of the SFS under the condition $\mathcal{C}_n$

Suppose the probability that j samples out of n are mutants at a polymorphic site is p_j . We may then represent the SFS using a generating function as $\sum_{j=1}^{n-1} p_j x^j$. The generating function representation will be used presently to define the extension operator and to derive a differential equation for the SFS.

Suppose the generating function is denoted S and that $n+1$ samples are related to the n samples via a binary merger. The probability that j out of $n+1$ samples are mutants is given by

$$p'_j = p_j \left(1 - \frac{j}{n}\right) + p_{j-1} \frac{j-1}{n}$$

for $j = 2, \dots, n-1$, by $p'_1 = p_1(1 - 1/n)$ for $j = 1$, and $p'_n = p_{n-1}(n-1)/n$ for $j = n$. Each of the n parents of the sample of size $n+1$ is equally likely to split into two. Therefore, to end up with j mutants among $n+1$ samples, we must have either j mutants among n samples with one of the non-mutants splitting (which occurs with probability $(1 - j/n)$) or $j-1$ mutants among n samples with one of the mutants splitting (which occurs with probability $(j-1)/n$). The extension operator $\mathcal{E}$ is defined by

$$\mathcal{E}S = \sum_{j=1}^n p'_j x^j.$$

The extension operator $\mathcal{E}$ applied to the SFS S results in the SFS of a sample of size greater by one that is assumed to be related to the parental sample via a single binary merger.

Suppose $k \in [k_{\min}(t), k_{\max}(t)]$. Under the condition $\mathcal{C}_k$ that those k ancestral samples coalesce (or exit the lens) with exactly one mutation, the k samples are hit with a mutation in the time interval $[t-dt, t]$ with a probability equal to

$$\frac{k\mu(t) dt q_0(k, t)}{q_1(k, t)}.$$

Thus, under the condition $\mathcal{C}_k$, the k samples are hit with a mutation at the rate

$$p(\mu|k, t) = \frac{k\mu(t)q_0(k, t)}{q_1(k, t)}. \quad (8)$$

Similarly, the k samples experience a binary merger with the conditional rate given by

$$p(\beta|k, t) = \frac{k(k-1)}{2N(t)} \frac{q_1(k-1, t)}{q_1(k, t)}. \quad (9)$$

It must be noted that $p(\mu|k, t)$ and $p(\beta|k, t)$ are rates and not probabilities.

Under the condition $\mathcal{C}_k$, there are three disjoint and exhaustive possibilities (with dt^2 terms ignored) for k ancestral samples in the time interval $[t-dt, t]$.

- The k ancestral samples are hit with a mutation with probability $p(\mu|k, t) dt$.
- The samples experience a binary merger with probability $p(\beta|k, t) dt$.
- There is no event with probability $(1 - p(\mu|k, t) - p(\beta|k, t)) dt$.

Let $S(k, t)$, with $t \in [0, T]$ and $k \in [k_{\min}(t), k_{\max}(t)]$, denote the SFS of k ancestral samples at time t under the condition $\mathcal{C}_k$. The three disjoint possibilities listed above imply that

$$S(k, t) = (1 - p(\mu|k, t) - p(\beta|k, t)) dt S(k, t-dt) + p(\mu|k, t) dt x + p(\beta|k, t) dt \mathcal{E}S(k-1, t-dt).$$

The middle term corresponds to a single mutant arising because of a mutation occurring in the interval $[t-dt, t]$. The last term corresponds to a binary merger.

In the limit $dt \rightarrow 0$, we obtain the differential equation

$$\frac{dS(k, t)}{dt} = -(p(\mu|k, t) + p(\beta|k, t))S(k, t) + p(\mu|k, t)x + p(\beta|k, t)\mathcal{E}S(k-1, t). \quad (10)$$

Fig. 3: Probability that a site is polymorphic as a result of a single mutation in the genealogy for demographic models 0, 1, 2 and various sample sizes. In all three plots, the black squares plot the probability for $n = 10^5$ samples that a site hit with a mutation has been hit with three or more mutations.

This differential equation is solved as an initial value problem from $t = 0$ to $t = T$. The SFS $S(k, 0)$ for $k \in [k_{min}(0), k_{max}(0)]$ is initialized using the exact formula when μ and N are both constant. For $k \notin [k_{min}(t), k_{max}(t)]$, $S(k, t) = 0$. In addition, $S(1, t) = 0$. The numerical solution of this differential equation is described in the Appendix. A computer program implementing the algorithm may be obtained from github.com/divakarvi/18-varymu.

Suppose $\mu(t) = \epsilon\nu(t)$ and we take $\epsilon \rightarrow 0$. The resulting limit is the zero mutation limit with varying mutation rate. An algorithm applicable to this limit is described in the Appendix.

SFS under the Wright-Fisher model

The Wright-Fisher model differs from the coalescent in that all mergers under the coalescent are binary mergers, whereas the Wright-Fisher model allows multiple mergers as well as simultaneous mergers [Durrett, 2008]. Harpak et al. [2016] considered whether the SFS computed assuming every merger to be a binary merger is reliable.

The formal truncation error in passing from the Wright-Fisher model to the coalescent is n^2/N [Durrett, 2008]. However, the leading truncation term in the SFS is only n/N and in fact the total variation distance in the SFS is only around 1% for $n = N = 20,000$ [Melfi and Viswanath, 2018b]. Furthermore, in demographic models such as 1 and 2 characterized by recent exponential growth, the total variation distance may be expected to be even lower. Assuming every merger to be a binary merger has little effect on the SFS.

Nevertheless, in the Appendix, we show how to compute the SFS with varying $\mu(t)$ and $N(t)$ under the Wright-Fisher model. An advantage of the Wright-Fisher model is that it is already discrete. We also remark that computations with the Wright-Fisher model may lend themselves to better optimization with the use of suitable asymptotic formulas.

Effect of Finite Mutation Rate on the SFS

Harpak et al. [2016] have presented evidence that the infinite sites model is violated for samples of around $n = 10^5$ haploid human genomes. The infinite sites models assumes that every new mutation in the genealogy occurs at a different site.

To understand when and how violations of the infinite sites model set in, we may first look

at the quantity

$$\frac{q_1(n, 0)}{1 - q_0(n, 0)},$$

which is the probability that there is exactly one mutation in the genealogy given that there are one or more mutations in the genealogy of the n samples. In Figure 3, we use $q_1/(1 - q_0)$ as a surrogate for the probability that a site is polymorphic as a result of a single mutation. The two probabilities will be close but are not exactly the same. They differ slightly because of the small probability that two mutations may occur in the same lineage in the genealogy and cancel each other. Thus, a sample with two or mutations in its genealogy may not be polymorphic.

In Figure 3, we have graphed the probability of a single mutation at a polymorphic site for demographic model 0 which assumes $N \equiv 2 \times 10^4$. Sample sizes of $n = 10^5$ would not make sense in the Wright-Fisher interpretation of that model. However, in the coalescent N is only a parameter to control rates of binary mergers.

The probability of a single mutation at a polymorphic site is the highest in demographic model 0. That appears to be because binary mergers are initially fastest in demographic model 0.

Both demographic models 1 and 2 assume exponentials that persist for more than a 100 generations. The population explosion slows down binary mergers and as a result the probability of double mutations is higher in demographic models 1 and 2. The exponential persists over a greater interval of time in model 1 and model 1 shows a greater probability of double mutations than model 2. In both models 1 and 2, the probability of a multiple mutation is noticeable for even $\mu = 10^{-8}$ and $n = 10^5$.

The probability that a site hit with a mutation has been hit with three or more mutations is given by

$$\frac{1 - q_0(n, 0) - q_1(n, 0) - q_2(n, 0)}{1 - q_0(n, 0)}.$$

For demographic model 1, $n = 10^5$, and $\mu = 10^{-7}$, the probability that a site that has been hit with a mutation has been hit with exactly one, exactly two, or three or more mutations is 60%, 28.4%, and 11.6%, respectively (see Figure 3). For demographic model 2, those probabilities are 78.3%, 18.4%, and 3.3%, respectively.

Harpak et al. [2016] have inferred a mutation rate of around 10^{-7} for CpG junctions. The SFS will be a mixture of the SFS due to a single mutation, the SFS due to two mutations, and the SFS due to three or more mutations in the genealogy. A sample being hit with three mutations at a site does not necessarily imply that the site shows four-fold polymorphism. Because transitions are more likely than transversions, the site may only be biallelic.

Suppose a sample carries multiple mutations at the same site in its genealogy. That does not imply that there is a haploid individual in the sample whose lineage has been hit with two mutations. The mutations may not be nested in the genealogical tree. If an ancestral lineage is hit with a mutation, the probability that the lineage is hit with yet another mutation may change [Nei, 1987].

In our analysis, we have assumed the probability to stay constant. Despite the limitations of that assumption, we may disentangle the effects of single and multiple mutations on the SFS. To do so, we turn to Figure 4. The total variation distance (or variation distance) between an

Fig. 4: The total variation distance between the SFS at a given μ and at $\mu = 0$ under the condition $\mathcal{C}_n$.

Fig. 5: The effect of finite μ on the SFS ($j < 20$).

SFS given by p_j and an SFS given by p'_j for $j = 1, \dots, n - 1$ is defined as

$$\frac{1}{2} \sum_{j=1}^{n-1} |p_j - p'_j|.$$

It is the right metric to use for comparison because it can be interpreted as the maximum difference between the probabilities of any possible event that is a subset of $\{1, \dots, n - 1\}$ [Brémaud, 1999]. Because it can be interpreted as a probability, the variation distance can also be thought of as a percent.

Figure 4 shows that the variation distance of the SFS with μ finite and $\mu = 0$ under the condition $\mathcal{C}_n$ falls as the sample size n increases. The variation distance is negligible for $\mu = 10^{-7}$. Thus, the phenomena described by Harpak et al. [2016] are entirely due to multiple mutations in the genealogy.

In addition to the variation distance of the SFS between μ finite and $\mu = 0$ being small, it is in a direction opposite to the total effect. Figure 5 shows that the effect of finite μ is to slightly increase the occurrence of rare alleles. However, in the overall SFS, rare alleles are depleted [Harpak et al., 2016]. Intuitively, we may understand why rare alleles are depleted by multiple mutations as follows. The probability of singletons ($j = 1$ mutants) begins to dominate in large samples as evident from Figure 5. When the genealogy carries two mutations, for example,

Fig. 6: The probability that a site is segregating (more precisely, hit with a mutation) as a function of the factor f (see (11)) for demographic models 1 and 2, respectively.

there is a substantial chance that both mutations are transitions. In large samples and under demographic models with exponentially growing population, both mutations are likely to hit near the leaves of the genealogical tree. The mutations are unlikely to be nested. Therefore, the probability of a singleton is depleted.

The Effect of Varying Mutation Rate on the SFS

We assume the mutation rate to be given by

$$\mu(\tau) = 10^{-9} \frac{T - \tau}{T} + \frac{10^{-9}}{f} \frac{\tau}{T}, \quad (11)$$

for $\tau \in [0, T]$, and $\mu(\tau) = 10^{-9}/f$ for $\tau \geq T$. The mutation rate is assumed to vary linearly in $[0, T]$. In addition, we assume $T = 367.8$ and $T = 119.47$ for demographic models 1 and 2, respectively. Because the historical variation of the mutation rate is only now being investigated, these assumptions are necessarily somewhat arbitrary. Genetic data is highly sensitive to μ , and there is no reason to assume anything faster than linear variation.

In the model for $\mu(t)$, the variation in μ sets in at $\tau = T$. For demographic models 1 and 2, we have assumed T to be to the epoch at which exponential increase in population sets in.

In the model, f is the factor by which the mutation rate increases from $\tau = T$, which is T generations in the past, to $\tau = 0$, which is the current time. We have assumed $\mu(t)$ to be 10^{-9} at the current time because that is the most widely quoted number, although it has been questioned recently [Gao et al., 2016, Scally and Durbin, 2012].

Methods used to infer the mutation rate rely directly or indirectly on the pioneering work of Watterson [1975] on the number of segregating sites. For a recent example, see [Bhaskar et al., 2015]. Therefore, it is appropriate to begin by looking at the probability $1 - q_0(n, 0)$ that a current sample of size n has been hit with a mutation at a site. The probability $1 - q_0(n, 0)$ is close to but not exactly the same as the probability that a site is segregating. If a site is hit with multiple mutations, there is a small probability that it is not segregating.

Figure 6 shows the way the number of segregating sites (more precisely, the probability a site is hit with a mutation) varies as a function of f . A sample size of $n = 100$ appears more

Fig. 7: Total variation distance of the SFS for a given f (see (11)) from the SFS with $f = 0$, which implies a constant mutation rate. The parameter f controls the variation in mutation rate.

Fig. 8: The effect of increasing and decreasing mutation rates on the SFS. The last plot uses demographic model 2 as well.

sensitive to variations in the mutation rate than larger samples, especially when $f > 1$ and $\mu(t)$ is increasing.

Figure 7 shows the total variation distance of the SFS between $f = f$ and $f = 0$ as a function of the parameter f . There is no easily discernible pattern in the plots. For the demographic model 1 and $n = 100$, the SFS changes by 6% and 3% when the mutation rate is assumed to double or halve, respectively. For demographic model 2, those numbers are 5% and 3%, respectively. The variation in $\mu(t)$ certainly has an effect on the SFS. Figure 8 shows that an increasing mutation rate augments the fraction of rare alleles, whereas a decreasing mutation rate depletes it.

Discussion

The Polanski-Kimmel algorithm for computing the SFS [Polanski and Kimmel, 2003] allows the population size $N(t)$ to vary but assumes the mutation rate to be constant and negligible. We have presented an algorithm to compute the SFS that allows the mutation rate $\mu(t)$ to be finite and varying. The main innovation in our algorithm is to ignore the internal structure of the genealogical tree and instead take a more Markovian approach.

The algorithm uses a first pass (with ancestral time τ increasing) to calculate the ancestral lens. The ancestral lens is then flipped to the calendar time lens. In the second pass, the SFS

as well as $q_0(k, t)$ and $q_1(k, t)$ are calculated with calendar time t increasing. The calculations solve a set of linear ordinary differential equations with time-varying coefficients.

The quantities $q_0(k, t)$ and $q_1(k, t)$ are the probabilities that a sample of size k coalesces without being hit by a mutation or after being hit by exactly one mutation, respectively. The algorithms for calculating q_0 and q_1 may be thought of as generalizations of the work of Watterson [1975] on the number of segregating sites. The generalization consists in allowing both $N(t)$ and $\mu(t)$ to vary arbitrarily.

When the sample size is $n \approx 10^5$ and the mutation rate is $\mu \approx 10^{-7}$, multiple mutations occur in human genealogies with appreciable probabilities. Algorithms for obtaining the SFS with two mutations have been applied to sample sizes of $n = 100$ [Jenkins et al., 2014]. In addition to reaching greater sample sizes, possibly while limiting calculations to rare alleles, there are yet other issues to be investigated. For example, once a site is hit with a mutation, the probability that it will be hit with yet another mutation will change. If the mutation probability is assumed to be constant, our calculations show that a polymorphic site may be hit with three mutations with a probability of around a few percent.

Appendix

In the appendix, we explain how the ancestral lens and the SFS are computed numerically. We also give a version of the algorithm that allows $\mu(t)$ to vary and then takes the zero mutation limit. In addition, we show how to calculate the SFS for the Wright-Fisher model with varying mutation rate and population size.

Implementation of the ancestral lens

The differential equation (5) is linear with non-constant coefficients. The decay rate $k(k-1)/2N(\tau)$ is around 10^4 for $k \approx 10^5$ and $N \approx 10^6$. Yet, the differential equations for $k = 2, \dots, n$ cannot be considered stiff because the decay rate corresponds to the rate of binary mergers and must be resolved. Out of precaution, we used a 4th order BDF discretization [Hairer et al., 1991]. There is no additional cost to using an implicit method because the equations are linear. Initial time-steps were taken using the implicit midpoint rule. The formal order of accuracy is 3.

Integrating the differential equations for $k = 2, \dots, n$ would be too expensive for large n . Instead, we restrict the integration to the ancestral lens even as it is being computed. Suppose the ancestral lens at τ is given by $[k_{\min}, k_{\max}]$. In the time step from τ to $\tau + h$, we use $k = \max(k_{\min} - 1, 1), \dots, k_{\max}$. After the time step, $k_{\min}(\tau + h)$ can be either $k_{\min}$ or $k_{\min} - 1$. It is equal to $k_{\min} - 1$ if $p(k_{\min} - 1, \tau + h) < 10^{-40}$ and $k_{\min}$ otherwise. Similarly, $k_{\max}(\tau + h)$ can be either $k_{\max} - 1$ or $k_{\max}$. It is equal to $k_{\max} - 1$ if $p(k_{\max}, \tau + h) < 10^{-40}$ and $k_{\max}$ otherwise.

Particular care is necessary during the very first time step. The ancestral lens at $\tau = 0$ is given by $k \in [n, n]$ and consists of a single point. If the above strategy is followed, the ancestral lens can grow to at most $k \in [n - 1, n]$ after the first time step. The tolerance of $\epsilon_{\text{lens}} = 10^{-40}$ is so small that the ancestral lens will in fact be much wider. Failing to capture its width correctly in the first step will corrupt the entire computation. To capture the width of the ancestral lens correctly, the implicit midpoint rule is iterated until the width of the ancestral

lens stabilizes. The way the ancestral lens grows during the very first step may be observed from Figure 2a.

Figure 2b shows an ancestral lens flipped to a calendar time lens. The flip is mostly straightforward except at the current epoch $t = T$ or $\tau = 0$. At $\tau = 0$, the ancestral lens is given by $k \in [n, n]$. However, $k \in [n, n]$ at $t = T$ will not do because the differential equations (6), (7), and (10) utilize information regarding $k - 1$ samples. When an implicit BDF discretization is involved, narrowing the lens suddenly at $t = T$ will create error in moving information from $n - 1$ samples to n . This problem is easily solved by taking the calendar time lens at t to be equal to the ancestral lens at $\tau = T - t + h$, where h is the time step into t .

Solution of the differential equations for q_0 , q_1 , and the SFS

Provided the calendar time lens is calculated carefully, no really new issues arise in the solution of the differential equations (6), (7), and (10) for $q_0(k, t)$, $q_1(k, t)$, and $S(k, t)$, respectively. The differential equations are discretized using 4th order BDF with initial time steps using the implicit midpoint rule. The differential equations for $q_0(k, t)$, $q_1(k, t)$, and $S(k, t)$ are solved simultaneously. Therefore, the memory requirements of this algorithm are very low.

The differential equations are solved only for $k \in [k_{min}(t), k_{max}(t)]$. During the time step from t to $t + h$, k_{min} or k_{max} (or both) may increase by 1. If q_0 is assumed to be 1 and q_1 , S are assumed to be zero outside the lens, as stated in the main text, no special handling is necessary when k_{max} increases. When k_{min} increases by 1, the values of $q_0(k_{min}, t)$ as well as $q_0(k_{min}, \cdot)$ at previous epochs that contribute to the step to $t + h$ are taken to be 1. Similarly, the corresponding values of q_1 and S are taken to be 0.

Choice of time step and accuracy

Suppose $dx/dt = -\alpha(t)x$. Then

$$\frac{d^4 x}{dt^4} = (\alpha(t)^4 - 6\alpha(t)^2 \dot{\alpha}(t) + 3\dot{\alpha}(t)^2 + 4\alpha(t)\ddot{\alpha}(t) - \ddot{\alpha}(t))x.$$

Because the numerical discretizations are formally of order 3, the time step h is obtained from the requirement

$$\left| \alpha(t)^4 - 6\alpha(t)^2 \dot{\alpha}(t) + 3\dot{\alpha}(t)^2 + 4\alpha(t)\ddot{\alpha}(t) - \ddot{\alpha}(t) \right| h^4 \leq htol.$$

In this requirement, we have taken $x = 1$ because all the differential equations are solving for probabilities. To preserve the numerical stability of the BDF formula, each new time step must be within a factor of 1.2 of the previous time step.

In the computation of the ancestral lens, we take

$$\alpha(\tau) = -\frac{k(k-1)}{2N(\tau)}$$

with $k = k_{max}(\tau)$. In the computation of $q_0(k, t)$, $q_1(k, t)$, and $S(k, t)$, we take

$$\alpha(t) = \frac{k(k-1)}{2N(t)} + k\mu(t)$$

with $k = k_{max}(t)$. Derivatives of α are computed using differences.

The accuracy of our program has been checked by comparing with implementations of Polanski-Kimmel algorithm [Bhaskar et al., 2014, 2015] and against coalescent simulations [Hudson, 2002]. In addition, we wrote an independent program in Python that solves the differential equations for $k = 1, \dots, n$ without limiting itself to an ancestral lens. The accuracy of the C program has been checked against the Python program. The Python program used `odeint()`, which is defined in the `scipy` library. All the reported computations have at least 4 digits of accuracy and often more than 10 digits of accuracy.

Zero limit with varying mutation rate

Our algorithm to compute the SFS simplifies slightly if we take $\mu(t) = \epsilon\nu(t)$ and then take $\epsilon \rightarrow 0$. The probability $q_0(k, t)$ that k ancestral samples at time t coalesce without being hit by a mutation is then $q_0(k, t) = 1 + \mathcal{O}(\epsilon)$. In the limit $\epsilon \rightarrow 0$, we have $q_0(k, t) \equiv 1$.

If we replace $q_1(k, t)$ by $\epsilon q_1(k, t)$ in (7) and take $\epsilon \rightarrow 0$, we obtain

$$\frac{dq_1(k, t)}{dt} = -\frac{k(k-1)}{2N(t)}q_1(k, t) + \nu(t)k + \frac{k(k-1)}{2N(t)}q_1(k-1, t).$$

Equations (8), (9), and (10) may still be used to compute the SFS, the only change being the replacement of $\mu(t)$ in (8) by $\nu(t)$.

The Wright-Fisher SFS with varying mutation rate and population size

Suppose the population size in the Wright-Fisher model is N_t at time t and suppose the mutation rate to be μ_t when the t -th generation begets the $t+1$ st generation. The probability that k samples at time $t+1$ have j parents is given by

$$p_{k,j,N_t} = \left\{ \begin{matrix} k \\ j \end{matrix} \right\} \left(1 - \frac{1}{N_t}\right) \cdots \left(1 - \frac{j-1}{N_t}\right) \frac{1}{N_t^{k-j}}.$$

Here $\left\{ \begin{matrix} k \\ j \end{matrix} \right\}$ is a Stirling number of the second kind.

Suppose $q_0(k, t)$ is the probability that k samples in generation t coalesce without being hit by a mutation. Then

$$q_0(k, t+1) = (1 - \mu_t)^k \sum_{j=1}^k p_{k,j,N_t} q_0(j, t).$$

Suppose $q_1(k, t)$ is the probability that k samples in generation t coalesce with exactly one mutation in their Wright-Fisher genealogy. Then

$$q_1(k, t+1) = k\mu_t(1 - \mu_t)^{k-1} \sum_{j=1}^k p_{k,j,N_t} q_0(j, t) + (1 - \mu_t)^k \sum_{j=1}^k p_{k,j,N_t} q_1(j, t).$$

Suppose $\mathcal{C}_k$ is the condition or event that k samples coalesce with exactly one mutation in their Wright-Fisher genealogy. Under the condition $\mathcal{C}_k$, the probability that a sample of size k is hit with a mutation during generation $t+1$ is

$$p(\mu|k, t+1) = \frac{k\mu_t(1 - \mu_t)^{k-1} \sum_{j=1}^k p_{k,j,N_t} q_0(j, t)}{q_1(k, t+1)}.$$

Similarly, the conditional probability that it has j parents in generation t but does not experience a mutation during generation $t + 1$ is

$$p(j|k, t + 1) = \frac{(1 - \mu_t)^k p_{k,j,N_t} q_1(j, t)}{q_1(k, t + 1)}.$$

The recurrence for the SFS is then given by

$$\mathcal{S}(k, t + 1) = p(\mu|k, t + 1) x + \sum_{j=1}^k p(j|k, t + 1) \mathcal{E}_k \mathcal{S}(j, t).$$

Here $\mathcal{E}_k \mathcal{S}(j, t)$ is an extension of the SFS for j samples to k samples, assuming the k samples to be children of the j samples. A formula for $\mathcal{E}_k$ follows from the Wright-Fisher propagators derived in [Melfi and Viswanath, 2018b].

The Stirling numbers are the main source of difficulty in implementing the Wright-Fisher model. However, uniform asymptotic formulas [Temme, 1993] can be used to make the Wright-Fisher implementation much more efficient. The coalescent is not much more than an asymptotic approximation to Stirling numbers of the second kind. If k children must be divided between j parents, under Wright-Fisher, the children are first partitioned into j sets in one of $\left\{ \begin{smallmatrix} k \\ j \end{smallmatrix} \right\}$ ways and each partition is assigned to a parent. The coalescent produces a partition of k children between j parents using binary mergers. If $k - j = 1$, the two partition distributions are the same. The approximation is not a particularly good one in general, when $k - j \gg 1$, although it works quite well for the SFS [Melfi and Viswanath, 2018b].

References

- V. Aggarwala and B.F. Voight. An expanded sequence context model broadly explains variability in polymorphism levels across the human genome. *Nature Genetics*, 48(4):349–355, 2016.
- G. Amster and G. Sella. Life history effects on the molecular clock of autosomes and sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(6):1588–1593, 2016.
- A. Bhaskar, A. G. Clark, and Y. S. Song. Distortion of genealogical properties when the sample is very large. *Proceedings of the National Academy of Sciences*, 111:2385–2390, 2014.
- A. Bhaskar, Y.X.R. Wang, and Y.S. Song. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Research*, 25(2):268–279, 2015.
- P. Brémaud. *Markov Chains*. Springer, 1999.
- J. Carlson et al. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *bioRxiv.org*, 2018.
- R. Durrett. *Probability models for DNA sequence evolution*. Springer Science & Business Media, 2008.

- M. Ehrlich and R.Y.-H. Wang. 5-methylcytosine in eukaryotic DNA. *Science*, 212(4501): 1350–1357, 1981.
- A. Eyre-Walker and Y.C. Eyre-Walker. How much of the variation in the mutation rate along the human genome can be explained? *G3: Genes, Genomes, Genetics*, 4(9):1667–1670, 2014.
- Z. Gao, M.J. Wyman, G. Sella, and M. Przeworski. Interpreting the dependence of mutation rates on age and time. *PLoS Biology*, 14(1), 2016.
- R.C. Griffiths. The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theoretical Population Biology*, 64(2):241–251, 2003.
- R.C. Griffiths and S. Tavaré. The age of a mutation in a general coalescent tree. *Stoch. Models*, 14:273–295, 1998.
- R.N. Gutenkunst, R.D. Hernandez, S.H. Williamson, and C.D. Bustamante. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, 2009.
- E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I*. Springer-Verlag, Berlin, 1991.
- J.B.S. Haldane. The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. *Annals of Eugenics*, pages 262–271, 1947.
- A. Harpak, A. Bhaskar, and J.K. Pritchard. Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans. *PLoS Genetics*, 12, 2016.
- K. Harris. Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11): 3439–3444, 2015.
- A. Hodgkinson and A. Eyre-Walker. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics*, 12:756–766, 2011.
- A. Hodgkinson, E. Ladoukakis, and A. Eyre-Walker. Cryptic variation in the human mutation rate. *PLoS Biology*, 7:0226–0232, 2009.
- R.R. Hudson. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18:337–338, 2002.
- D.G. Hwang and P. Green. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 101(39):13994–14001, 2004.
- P.A. Jenkins, J.W. Mueller, and Y.S. Song. General triallelic frequency spectrum under demographic models with variable population size. *Genetics*, 196(1):295–311, 2014.
- P.L.F. Johnson and I. Hellmann. Mutation rate distribution inferred from coincident SNPs and coincident substitutions. *Genome Biology and Evolution*, 3(1):842–850, 2011.

- S.-H. Kim, N. Elango, C. Warden, E. Vigoda, and S.V. Yi. Heterogeneous genomic molecular clocks in primates. *PLoS Genetics*, 2(10):1527–1534, 2006.
- I. Mathieson and D. Reich. Differences in the rare variant spectrum among human populations. *PLoS Genetics*, 13(2), 2017.
- A. Melfi and D. Viswanath. Single and simultaneous binary mergers in Wright-Fisher genealogies. *Theoretical Population Biology*, 121:60–71, 2018a.
- A. Melfi and D. Viswanath. The Wright-Fisher site frequency spectrum as a perturbation of the coalescent's. *bioRxiv.org*, 2018b.
- J.J. Michaelson, Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, W. Wu, R. Corominas, A. Peoples, A. Koren, A. Gore, S. Kang, G.N. Lin, J. Estabillio, T. Gadomski, B. Singh, K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L.M. Iakoucheva, Y. Li, J. Wang, and J. Sebat. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell*, 151(7):1431–1442, 2012.
- P. Moorjani, C.E.G. Amorim, P.F. Arndt, and M. Przeworski. Variation in the molecular clock of primates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(38):10607–10612, 2016a.
- P. Moorjani, Z. Gao, and M. Przeworski. Human germline mutation and the erratic evolutionary clock. *PLoS Biology*, 14(10), 2016b.
- V.M. Narasimhan, R. Rahbari, A. Scally, A. Wuster, D. Mason, Y. Xue, J. Wright, R.C. Trembath, E.R. Maher, D.A. Van Heel, A. Auton, M.E. Hurles, C. Tyler-Smith, and R. Durbin. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nature Communications*, 8(1), 2017.
- M. Nei. *Molecular Evolutionary Genetics*. Columbia University Press, 1987.
- M.R. Nelson, D. Wegmann, M.G. Ehm, D. Kessner, P. St. Jean, C. Verzilli, J. Shen, Z. Tang, S.-A. Bacanu, D. Fraser, L. Warren, J. Aponte, M. Zawistowski, X. Liu, H. Zhang, Y. Zhang, J. Li, Y. Li, L. Li, P. Woollard, S. Topp, M.D. Hall, K. Nangle, J. Wang, G. Abecasis, L.R. Cardon, S. Zöllner, J.C. Whittaker, S.L. Chisoe, J. Novembre, and V. Mooser. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, 336(6090):100–104, 2012.
- A. Polanski and M. Kimmel. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics*, 165:427–436, 2003.
- A. Polanski, A. Bobrowski, and M. Kimmel. A note on distributions of times to coalescence, under time-dependent population size. *Theoretical Population Biology*, 63(1):33–40, 2003.
- R. Rahbari, A. Wuster, S.J. Lindsay, R.J. Hardwick, L.B. Alexandrov, S. Al Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, M.R. Stratton, and M.E. Hurles. Timing, rates and spectra of human germline mutation. *Nature Genetics*, 48(2):126–133, 2016.

-
- S.A. Sawyer and D.L. Hartl. Population genetics of polymorphism and divergence. *Genetics*, 132(4):1161–1176, 1992.
- A. Scally and R. Durbin. Revising the human mutation rate: Implications for understanding human evolution. *Nature Reviews Genetics*, 13(10):745–753, 2012.
- M. Steinrücken, E. M. Jewett, and Y. S. Song. SpectralTDF: transition densities of diffusion processes with time-varying selection parameters, mutation rates and effective population sizes. *Bioinformatics*, 32(5):795–797, 2016.
- N.M. Temme. Asymptotic estimates of Stirling numbers. *Studies in Appl. Math.*, 89:233–243, 1993.
- D. Živković, M. Steinrücken, Y.S. Song, and W. Stephan. Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics*, 200(2):601–617, 2015.
- G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7:256–276, 1975.