Available online at www.sciencedirect.com
ScienceDirect
Theoretical Population Biology 73 (2008) 342–348
Theoretical Population Biology
www.elsevier.com/locate/tpb

Can one learn history from the allelic spectrum?

Simon Myers[a], Charles Fefferman[b], Nick Patterson[a,*]

[a] Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142, United States
[b] Deptartment of Mathematics, Fine Hall, Washington Road, Princeton, NJ 08544, United States

Received 17 March 2007
Available online 30 January 2008

Abstract

It is well known that the neutral allelic frequency spectrum of a population is affected by the history of population size. A number of authors have used this fact to infer history given observed allele frequency data. We ask whether perfect information concerning the spectrum allows precise recovery of the history, and with an explicit example show that the answer is in the negative. This implies some limitations on how informative allelic spectra can be.

Keywords: Diffusions; Allelic spectrum; Kimura

1. Introduction

It is well known that the neutral allelic frequency spectrum of a population is affected by the population history. A number of authors have exploited this fact so as to infer population history given observed allele frequency data (Przeworski et al., 2000; Wakeley et al., 2001; Marth et al., 2004; Adams and Hudson, 2004; Nielsen, 2004; Williamson et al., 2005; Schaffner et al., 2005; Voight et al., 2005; Garrigan and Hammer, 2006; Chen et al., 2007). This raises questions about the relationship between ancient population size changes, and the frequency spectrum at mutant sites. In particular we might ask how much information can be obtained about ancestral demography of a population by typing many unlinked markers in a sample of individuals.

As more samples are gathered, in principle it is possible to gain increasingly precise information about the frequency spectrum. Therefore, we here examine two related questions regarding the history of a population. First, how much information can one obtain about population history given a sample of n individuals? Second, suppose we knew the frequency spectrum *exactly*. Is the history of past population sizes then completely determined?

Throughout this paper, we make two simplifying assumptions. First, we assume that the population is panmictic and large enough to allow drift to be approximated by a diffusion process. Second, we assume that mutation is rare in the population so that the infinite sites model is appropriate, and we can regard every segregating site as having arisen from a unique mutation event. We consider the frequency spectrum f of the derived (mutant) allele, where $f(y)$ is defined as the expected frequency of mutations of population frequency y within the genome. All mutations and alleles in this paper are considered to be selectively neutral, so 'frequency spectrum' would more exactly be described as 'neutral frequency spectrum'. We will regard the frequency spectrum as only defined up to an arbitrary normalizing constant. This implies that the mutation rate (assumed constant over time) is not relevant.

We ask, given the frequency spectrum f, is the history of past population sizes determined? This question turns out to be formally similar to the famous question *Can one hear the shape of a drum?* (Gordon et al., 1992), and like that question the answer is in the negative.

2. Preliminary theory

Consider a panmictic population, assumed for now to be of constant large size N. Under several different population genetic models, for example the discrete time Wright–Fisher model, and the continuous time Moran model, the frequency

* Corresponding author.
E-mail address: nickp@broad.mit.edu (N. Patterson).

0040-5809/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2008.01.001

dynamics of a mutation in the population are often modelled using a well known diffusion process. In this setting, it is standard to measure time in units of $2N$ generations. For example, the drift in frequency of some mutation over k (discrete time) generations in the Wright–Fisher model corresponds approximately to drift over $k/2N$ (continuous time) units in the standard diffusion approximation. With this usual diffusion approximation, define the *Kimura transition function* $K(x, y; \tau)$ to be the probability density that the allele frequency now (at time 0) is y given that the allele frequency at time $-\tau$ was x. We assume here that $0 < x, y < 1$. A theorem of Kimura (1955) proves

$$K(x, y; \tau) = x(1-x)\sum_{i=0}^{\infty} \frac{J_i(x)J_i(y)}{X_i} e^{-\lambda(i)\tau} \tag{1}$$

with $\lambda(i)$ given by:

$$\lambda(i) = \frac{(i+1)(i+2)}{2} \tag{2}$$

where J_i are explicit polynomials (Jacobi polynomials) orthogonal under the weight function $x(1-x)$ and X_i is a normalization constant so that

$$\int_0^1 x(1-x)J_i(x)J_j(x)dx = \delta_{ij}X_i \tag{3}$$

and δ_{ij} is the Kronecker delta. Explicit formulae for J_i and X_i are given in the appendix of Patterson (2005) (where X_i is written as N_i). Here our J_i are related to the Jacobi polynomials $P_n^{(1,1)}$ (defined on $[-1, 1]$) by

$$J_i(x) = P_i^{(1,1)}(2x-1).$$

Write

$$K(x, y; \tau) = x(1-x)K_0(x, y; \tau)$$

and

$$\lim_{x\to 0} K_0(x, y; \tau) = Q(y, \tau).$$

Then by Taylor's theorem,

$$K(\epsilon, y; \tau) = \epsilon Q(y, \tau) + O(\epsilon^2). \tag{4}$$

From Eq. (1)

$$Q(y, \tau) = \sum_{i=0}^{\infty} c_i J_i(y) e^{-\lambda(i)\tau} \tag{5}$$

where

$$c_i = J_i(0)/X_i. \tag{6}$$

We are interested in the more general setting where the ancestral population size changes through time. Once again, the Wright–Fisher process can be approximated by a diffusion. It is well understood that by choosing the appropriate transformation of time, from time in generations to the diffusion timescale, this diffusion reduces to the constant-size case.

Suppose that the population size t generations ago was $N(t)$, and define $D(x, y; t)$ to be the conditional probability of y given that the frequency t generations ago was x. To simplify subsequent equations, we abuse notation slightly by allowing t, the time in generations, to be continuous — this can be naturally achieved, for example, by defining $N(t)$ at noninteger time points using a step function $N(t) = N(\lfloor t \rfloor)$ and then taking $D(x, y, t) = D(x, y, \lfloor t \rfloor)$. We will assume that $N(t)$ is bounded away from 0 and also bounded above. So there exist $N_{\min}, N_{\max}$ with

$$N_{\min} \le N(t) \le N_{\max}$$

for all t. We set $N = N_{\min}$. Note that t is still 'calendar time', clocked by the number of generations.

To transform to the diffusion timescale, define

$$\tau(t) = \int_0^t \frac{1}{2N(s)} ds. \tag{7}$$

It can be shown that the following equation is approximately true for large N:

$$D(x, y; t) = K(x, y; \tau(t)). \tag{8}$$

This approximation becomes exact in the limit as $N \to \infty$, in the following sense. Suppose that after a linear rescaling of time, the (relative) population size changes remain fixed as N increases, so that we may write $N(t) = NG(t/2N)$ for some bounded function G. Now setting $t = 2N\alpha$, where α remains constant, gives

$$\tau(t; N) = \int_0^t \frac{1}{2N(s)} ds = \int_0^{\alpha} \frac{1}{G(v)} dv$$

so τ is independent of N. Finally, allowing $N \to \infty$, $D(x, y; t) \to K(x, y; \tau)$. Eq. (8) is implicit in Slatkin and Hudson (1991, Equation (4)) and explicit in Griffiths and Tavaré (1994, Equation (3)). Throughout the remainder of this paper, we assume that N is large enough that Eq. (8) can be considered an identity. Note that

$$\frac{\partial \tau}{\partial t} = \frac{1}{2N(t)}. \tag{9}$$

For the simple case that $N(t)$ is the constant N then Eq. (7) shows that τ measures time in units of $2N$ generations, the standard normalization.

Considering time rescaled in this way, it is possible to obtain the frequency spectrum at segregating sites. We take an approach used by Kimura and Maruyama (1975), clarified in Watterson (1976) and Sawyer (1977). This is discussed in detail in Patterson (2005).

For a constant size population, take a time T, much larger than N generations, where N is the population size. In this case, mutations occur in the population at a constant rate per generation, and when a new mutation occurs it has initial frequency $1/2N$. Now and later, we will assume a unit mutation rate in the diffusion time scale for definiteness. Then the density of alleles in the population at frequency y is proportional to

$$\int_0^T D(1/2N, y, t)dt.$$

Letting $T \to \infty$, and transforming to units of genetic time, the frequency spectrum $P(y)$ of the derived allele is

$$P(y) \propto \int_0^\infty 2NK(1/2N, y, \tau)d\tau = \int_0^\infty Q(y, \tau)d\tau + O(1/N).$$

Taking the limit as $N \to \infty$, Ewens (1963) showed that

$$\int_0^\infty Q(y, t)dt = \frac{2}{y}. \tag{10}$$

We now generalize this argument to the case of a changing population size. If the population size t generations ago is $N(t)$ then the rate at which mutations occur in the population at time t is proportional to $N(t)$, and if a mutation occurs, then the allele has frequency $1/2N(t)$.

Since $N(t)$ is of order $N = N_{\min}$, which is large, we will assume that under our diffusion approximation we only need to retain the term of lowest order in $1/N(t)$. Thus

$$P(y) \propto \int_0^\infty \frac{N(t)}{N} D(1/2N(t), y; t)dt.$$

We will change variables from t to τ using Eq. (7).

It is convenient to define

$$\tilde{N}(\tau(t)) = N(t). \tag{11}$$

Also, using Eq. (9) we have

$$t(\tau) = \int_0^\tau 2N(\tau)d\tau$$

and so

$$N(t(\tau)) = \tilde{N}(\tau). \tag{12}$$

Eqs. (11) and (12) show a $(1, 1)$ correspondence between functions N and $\tilde{N}$. We describe time indexed by τ as 'genetic time' in which population frequencies diffuse at a constant rate. This is in contrast to calendar time, indexed by generations.

Let $n(t) = N(t)/N$, and $\tilde{n}(\tau) = n(t(\tau))$. Set

$$P^\star(y) = \int_0^\infty n(t)D(1/2N(t), y; t)dt.$$

Retaining terms of lowest order in $1/N$, then the frequency spectrum $P(y)$ has

$$P(y) \propto P^\star(y)$$

where

$$\begin{aligned} P^\star(y) &= \int_0^\infty n(t)D(1/2N(t), y; t)dt \\ &= 2\int_0^\infty \tilde{n}(\tau)\tilde{N}(\tau)K(1/2\tilde{N}(\tau), y; \tau)d\tau \\ &= \int_0^\infty \tilde{n}(\tau)Q(y, \tau)d\tau + O(1/N) \\ &\approx \int_0^\infty \tilde{n}(\tau)Q(y, \tau)d\tau. \end{aligned} \tag{13}$$

From this last expression, and using (10) we see that to order $1/N$

$$2 \le yP^\star(y) \le 2\frac{N_{\max}}{N_{\min}}$$

so that we have indeed retained the leading term of $P^\star(y)$.

Eq. (13) which is Eq. (52) of Griffiths (2003) in a different notation, generalizes Eq. (10).

Now, using Eq. (5) we find:

$$P(y) \propto \sum_{i=0}^\infty c_i J_i(y) \int_0^\infty \tilde{n}(\tau)e^{-\lambda(i)\tau}d\tau. \tag{14}$$

In this paper we will only consider the frequency spectrum as scaled by an arbitrary constant (that is we do not consider the overall polymorphism rate). $P(y)$ is an improper distribution (with a simple pole at 0). We choose to scale P so that

$$\int_0^1 y(1-y)P(y)dy = 1/6. \tag{15}$$

It follows from Eqs. (14) and (15) that there exists some constant C so that

$$P(y) = \sum_{i=0}^\infty d_i J_i(y) \tag{16}$$

where

$$d_i = Cc_i \int_0^\infty \tilde{n}(\tau)e^{-\lambda(i)\tau}d\tau.$$

Further, from the orthogonality of the $J_i(y)$, Eq. (15) implies that

$$\int_0^1 d_0 y(1-y)J_0(y)dy = 1/6.$$

As $J_0(y) = 1$, it follows that $d_0 = 1$. So d_i may be computed from $\tilde{n}$ by setting

$$d'_i = c_i \int_0^\infty \tilde{n}(\tau)e^{-\lambda(i)\tau}d\tau$$

and then setting

$$d_i = \frac{d'_i}{d'_0}.$$

Eq. (16) is the key to much of what follows and demonstrates that even when the population size varies, the frequency spectrum can be written as a sum of 'harmonics' (in this case Jacobi polynomials) that decay with time, the high order harmonics decaying rapidly. This reduction of the effect of a given population demography to a countable set of coefficients d_i immediately implies that any two demographies producing the same coefficients must result in the same population frequency spectrum.

Let us suppose that we have ascertained K unlinked polymorphic biallelic markers (such as SNPs) and observe allele counts (a_k, b_k) for marker k where a_k is the count for the derived allele. We can assume from the ascertainment, that

$a_k, b_k > 0$. Then the likelihood of the observations for marker k is

$$\mathcal{L}(k) = \int_0^1 y^{a_k}(1-y)^{b_k} P(y)dy \tag{17}$$

$$= \sum_{i=0}^{\infty} d_i \int_0^1 y(1-y)y^{a_k-1}(1-y)^{b_k-1} J_i(y)dy. \tag{18}$$

Properties of the Jacobi polynomials show that all terms of this sum vanish for $i > (a_i + b_i - 2)$. Define

$$X(a,b,i) = \int_0^1 y^a (1-y)^b J_i(y)dy$$

then the log likelihood for our observations is given by

$$\mathcal{L} = \sum_k \log \sum_{i=0}^{a_k+b_k-2} d_i X(a_k, b_k, i). \tag{19}$$

3. Consequences

Eq. (19) may be interesting in analysing SNP data for information about demographic history. Here we simply remark that by the likelihood principle all information concerning $N(t)$ available from the data is encoded in the coefficients d_i of Eq. (19). Further, even as the number of observed mutations tends to infinity, any real sample of finite size n only provides information about the first n coefficients $d_1, d_2, \ldots d_n$. In practice many fewer such d_i are likely to be usefully estimated, because the higher order 'harmonics' decay rapidly. [Here the sample size n may be taken as the largest size available at any locus.]

Now suppose that we know all the d_i *exactly*, an unrealistically favourable assumption. A natural question to ask is: *does this determine the function* $\tilde{n}(\tau)$? In the remainder of the paper we show the answer is negative.

Since c_i is independent of the data, we can answer our question by exhibiting a nonzero function $D(\tau)$ on $[0, \infty)$ that is bounded and for which

$$\int_0^{\infty} D(\tau)e^{-\lambda(i)\tau}d\tau = 0 \tag{20}$$

for every $i = 0, 1, \ldots$. To avoid trivialities we require that

$$\int_0^{\infty} D^2(\tau)d\tau > 0.$$

Then if we take $\tilde{N}(\tau)$ to be a constant N a second function with the same frequency spectrum will be a demography $X(t)$ such that $\tilde{X}(\tau) = N + \alpha D(\tau)$ where we choose α small enough so that $\tilde{X}(\tau)$ is positive everywhere.

4. Müntz–Szasz theory

Thus we seek a nonzero function $D(\tau)$ such that

$$\int_0^{\infty} D(\tau)e^{-\frac{k(k+1)\tau}{2}}d\tau = 0$$

for every $k = 1, 2, \ldots$. Change variables by $u = e^{-\tau}$. Then we need a nonzero function $f(u) = D(-\log u)$ such that

$$\int_0^1 f(u)u^{\frac{k(k+1)}{2}-1}du = 0$$

for every $k = 1, 2, \ldots$.

Given a set E of nonnegative exponents e_i $(i = 1, 2, \ldots)$, then it follows from the Müntz–Szasz theorem (see for example DeVore and Lorentz (1993, Chapter 11)) that a necessary and sufficient condition that there exists a square-integrable f with $\int_0^1 f^2(u)du > 0$ and

$$\int_0^1 f(u)u^e du = 0$$

for every $e \in E$ is that $\sum_{e\in E, e\neq 0} 1/e$ converges. This generalizes the Weierstrass approximation theorem which in effect takes E to be the set of positive integers, and states that any continuous function on the unit interval can be approximated arbitrarily well by polynomials. In our case

$$E = \{k(k+1)/2 - 1\}, \quad k = 1, 2, \ldots.$$

Then $\sum_{e\in E, e\neq 0} 1/e$ converges, and hence a function f satisfying the required conditions can be found.

Indeed choose an integer $m > 0 \notin E$. If h is an integrable function on [0, 1] set $\|h\|$ (the L_2 norm) to be

$$\|h\| = \left(\int_0^1 h^2(x)dx\right)^{\frac{1}{2}}.$$

Now define $f_n = \sum_{k=1}^n c_k x^{k(k+1)/2-1}$ so as to minimize $\|x^m - f_n\|$. We can show:

1. $\lim_{n\to\infty} f_n(x) \to f(x)\ \forall x \in [0, 1]$.
2. $f(x)$ is continuous, bounded and infinitely differentiable in $[0, 1)$.
3. $g(x) = x^m - f(x)$ has a norm greater than 0.
4. $\int_0^1 g(x)x^e dx = 0$ for every $e \in E$.

(Numerical evidence strongly suggests that f is in fact well-behaved also at $x = 1$, but the convergence is delicate and we do not claim a proof. If necessary we can modify f by a suitable convolution, forming $f^\star$, using an argument given below, and take g to be $x^m - f^\star$). Thus we can take $D = g$ to be the function that we seek. However, the details of our argument are complicated, and we prefer to give a simpler example.

5. An explicit example

Let

$$f_1(t) = \frac{\cos(\pi^2/t)\exp(-t/8)}{\sqrt{t}}.$$

Then the Laplace transform

$$\hat{f}_1(s) = \int_0^{\infty} f_1(t)e^{-st}dt \quad (s > 0)$$

$$= \cos\left(\frac{\pi}{2}\sqrt{(1+8s)}\right)A(s) \tag{21}$$

where:

$$A(s) = \frac{2^{3/2}\sqrt{\pi}\exp\left(-\frac{\sqrt{(1+8s)}\pi}{2}\right)}{\sqrt{1+8s}}. \tag{22}$$

$A(s) > 0$ for every $s \geq 0$. The integral is not an easy one, and in the appendix we prove

$$\int_0^\infty \frac{e^{-t}\cos(a^2/t)}{t^{\frac{1}{2}}}dt = \sqrt{\pi}\exp(-a\sqrt{2})\cos(a\sqrt{2}) \tag{23}$$

from which Eqs. (21) and (22) follow readily. This is also an entry in standard tables (Gradshteyn and Ryzhik, 1979, Equation 3.967(2)).

It follows that:

$$\int_0^\infty f_1(t)e^{-\lambda(i)t}dt = 0$$

for every $i = 0, 1, \ldots$. Evidently, $f_1(t)$ is badly behaved near $t = 0$. To fix this, define $f_0(t) = \exp(-1/t^2)$ and

$$F(t) = \int_0^t f_0(t-u)f_1(u)du. \tag{24}$$

Then by the convolution property of Laplace transforms

$$\int_0^\infty F(t)e^{-\lambda(i)t}dt = 0.$$

Define $F(0) = 0$. It is easy to check that F is continuous and indeed infinitely differentiable on the nonnegative reals.

By taking convolutions of arbitrary functions with $f_1(\tau)$ it is possible to generate a diverse family of bounded functions, each of which is orthogonal to the family $e^{-\lambda(i)\tau}$, $i = 0, 1, \ldots$. For any member D of this family (for instance taking $D(\tau) = F(\tau)$ above), we can add a small multiple of D to any population history $\tilde{H}$, (defined in genetic time) and produce an alternative history with the same allelic spectrum as $\tilde{H}$. As the map $\tilde{H} \to H$ is one-to-one this constructs families of population histories in calendar time all with the same spectrum.

[Explicitly define

$$k(u) = 1 + \alpha D(u)$$

and

$$h(t) = k(\tau(t)) = 1 + \alpha D(\tau(t))$$

Then

$$\int_0^{\tau(t)} (1 + \alpha D(y))dy = t$$

so that $\tau(t)$ is the inverse function of the integral of $1+\alpha D(y)$.].

Frequency spectrum data can therefore never fully specify population history.

6. Discussion and conclusion

We show in Fig. 1 a function $F(\tau)$ (see Eq. (24)) such that all integrals

$$\int_0^\infty F(\tau)\exp(-\lambda(i)\tau)d\tau = 0.$$

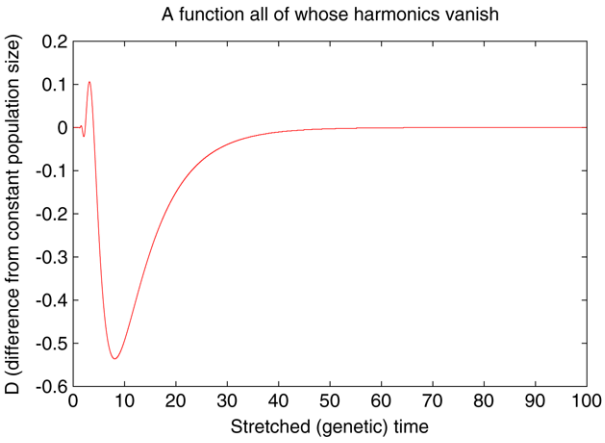

Fig. 1. We show a function $F(t)$ such that all integrals $\int_0^\infty F(\tau)\exp(-\lambda(i)\tau)d\tau = 0$.

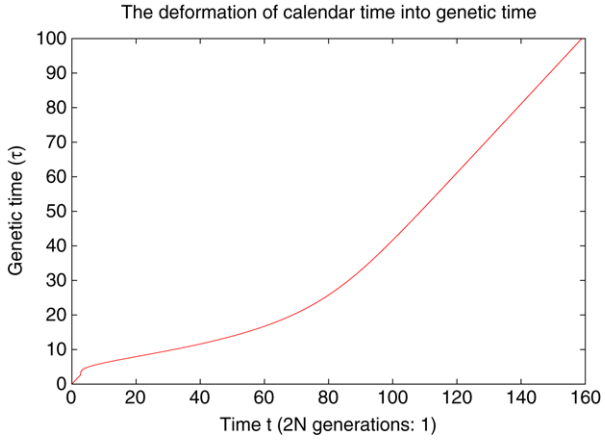

Fig. 2. Form $\tilde{N}(\tau) = 2N(1 - 9F(\tau))$ where F is the function of the previous figure, and N is some baseline population size. $\tilde{N}$ is always positive. We show genetic time (τ) as a function of calendar time t.

Given allele frequency data, it is in principle impossible to distinguish from allele frequency data a history $\tilde{N}(\tau)$ and an alternate history $\tilde{N}'(\tau) = \tilde{N}(\tau)+\alpha F(\tau)$ provided that for every τ, $\tilde{N}'(\tau) > 0$. As an example, choose a population size X and set $\tilde{N}(\tau) = X(1-9F(\tau))$. $\tilde{N}$ is positive, bounded and bounded away from zero. In Fig. 2 we plot genetic time τ as a function of calendar time t, where we use Eq. (7) for the calculations. Then we show in Fig. 3, the corresponding population size history as a function of calendar time. Informally, changes in population size at some past time are cancelled out by other changes in the opposite direction. The population size is reasonably constant at first, but there is a a strong bottleneck later that is completely invisible in the frequency spectrum. It is possible to construct alternative histories (not shown) with more recent 'invisible' bottlenecks or expansions.

By measuring the frequency spectrum, we gain information about the coefficients d_i. This is equivalent to measuring the projections of the population history onto the function space spanned by the basis functions $e^{-\lambda(i)\tau}$, $i = 0, 1, \ldots$.

We can never gain information about that component of population history which is orthogonal to this basis set, and which we have shown here to be nonzero in general. Further, in

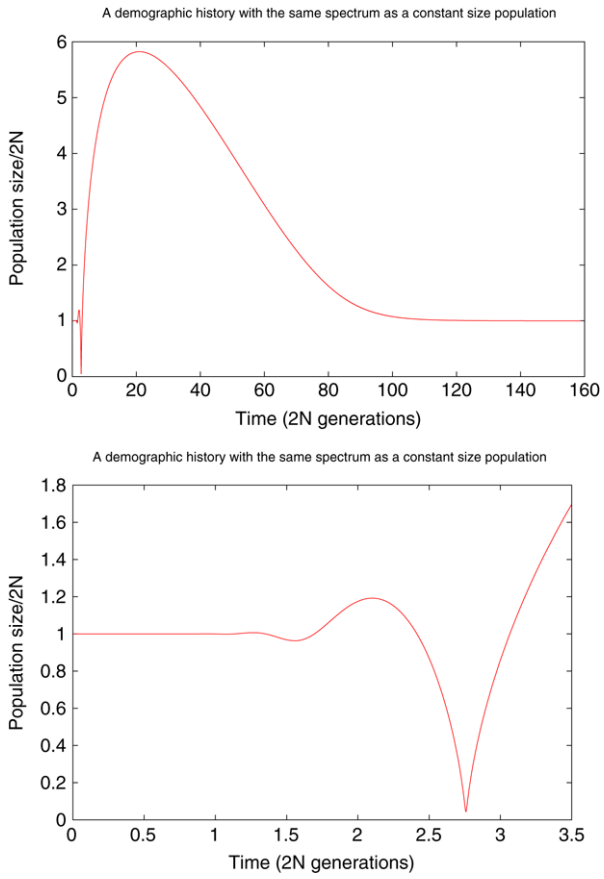

Fig. 3. In the top figure we show population size for a history corresponding to $\tilde{N}(\tau)$ of Fig. 2. Most of the interesting structure is for relatively small times, and so we also show an expansion of the figure for time $t \leq 3.5$.

practice we anticipate that only a limited number of coefficients $\{d_i\}$ are likely to strongly contribute to the observed spectrum. This suggests a natural way to perform inference about history based on the frequency spectrum, concentrating on that component of past ancestry about which the data provide information. We might consider only histories expressible in terms of an orthonormal basis (in the L_2 norm) of functions, constructed from the original basis above. Such an approach would be sensible in cases where limited prior knowledge about population history exists, or to enable inference that does not depend on specific assumptions about historical events.

Our frequency spectrum findings relate only to the use of unlinked neutral loci to infer population histories. Unlinked markers might at first appear to be most informative (because each marker contributes independent information), but in fact the correlation between linked loci provides additional information in inferring population size histories. Indeed, if it were possible to observe full genealogies at many loci rather than just SNP frequencies, we believe it would in theory be possible to accurately reconstruct such histories. Although we cannot directly observe a genealogy, this does suggest that utilizing joint variation patterns at groups of tightly linked markers will substantially improve ancestry inference, at the cost of introducing additional methodological challenges.

Appendix

Theorem 1. *Let* $a > 0$ *and*

$$X(a) = \int_0^\infty \frac{e^{-t}\cos(a^2/t)}{t^{\frac{1}{2}}}dt.$$

Then

$$X(a) = \sqrt{\pi}\exp(-a\sqrt{2})\cos(a\sqrt{2}).$$

Proof. We need the following lemma:

Lemma 1. *Let* $a > 0$, $u \geq 0$ *and define*

$$V(a,u) = \int_0^\infty \frac{e^{-t}}{t^{\frac{1}{2}}}\exp(-ua^2/t)\exp(ia^2/t)dt$$

so that $X(a)$ *is the real part of* $V(a,0)$. *Then*

$$V(a,u) = \sqrt{\pi}\exp(-2a\sqrt{u-i})$$

where on taking square roots we choose the root with positive real part.

Proof. Assume $u > 0$. Define for $s > 0$

$$\hat{V}(s,u) = \int_0^\infty a^{s-1}V(a,u)da. \tag{A.1}$$

This is the Mellin transform of V. We will need repeatedly the standard integral

$$\int_0^\infty x^{p-1}e^{-zx}dx = \frac{\Gamma(p)}{z^p} \tag{A.2}$$

valid for the real part of z being > 0, which is the integral yielding the characteristic function of the gamma distribution.

Now

$$\hat{V}(s,u) = \int_0^\infty a^{s-1}\int_0^\infty \frac{e^{-t}}{t^{\frac{1}{2}}}\exp(-ua^2/t)\exp(ia^2/t)dtda.$$

The integral is absolutely convergent and so by Fubini's theorem we can interchange the order of integration. Thus

$$\begin{aligned}\hat{V}(s,u) &= \int_0^\infty \frac{e^{-t}}{t^{\frac{1}{2}}}\int_0^\infty a^{s-1}\exp(-ua^2/t)\exp(a^2i/t)dadt\\ &= \int_0^\infty \frac{e^{-t}}{2t^{\frac{1}{2}}}\int_0^\infty b^{s/2-1}\exp(-ub/t)\exp(ib/t)dbdt\\ &= \int_0^\infty \frac{e^{-t}\Gamma(s/2)t^{s/2}}{t^{\frac{1}{2}}(u-i)^{s/2}}dt\\ &= \frac{\Gamma(s/2)\Gamma((s+1)/2)}{2(u-i)^{s/2}}\end{aligned}$$

where we apply Eq. (A.2) to evaluate the inner integral, then recognize the outer integral as the standard Gamma integral. Now we apply the Legendre duplication formula

$$\Gamma(2z) = \frac{1}{\sqrt{2\pi}}\Gamma(z)\Gamma(z+\frac{1}{2})2^{2z-\frac{1}{2}}$$

to obtain:

$$\hat{V}(s,u) = \sqrt{\pi}\frac{\Gamma(s)}{2^s(u-i)^{s/2}}.$$

Write

$$(u-i)^{s/2} = (\sqrt{u-i})^s$$

where we take the square root with positive real part. Then, applying Eq. (A.2) again, we see that

$$\hat{V}(s,u) = \sqrt{\pi}\int_0^\infty a^{s-1}\exp\left(-2a\sqrt{u-i}\right)da.$$

By the uniqueness of the Mellin transform, this is enough to prove our lemma for $u > 0$. Now let u tend to 0 from above. The integrand of $V(a,u)$ is dominated by $e^{-t}/t^{\frac{1}{2}}$ and thus $V(a,u) \to V(a,0)$. This shows that

$$\begin{aligned}\int_0^\infty \frac{e^{-t}}{t^{\frac{1}{2}}}\exp(ia^2/t)dt &= \sqrt{\pi}\exp(-2a\sqrt{-i}) \\ &= \sqrt{\pi}\exp\left(\frac{-2a(1-i)}{\sqrt{2}}\right) \\ &= \sqrt{\pi}\exp(-\sqrt{2}a)\exp(\sqrt{2}ai).\end{aligned}$$

Taking real parts this proves our theorem. □

A related, somewhat easier argument for the integral

$$\int_0^\infty \frac{e^{-(ut+x^2/4t)}}{t^{\frac{1}{2}}}dt$$

is given in Bellman (1961, Page 30).

[The introduction of u may seem an unnecessary complication, but at $u = 0$ the integral is only conditionally convergent, and it is not clear how to justify the change in the order of integration. This way we can interchange the integration order when $u > 0$, and then let $u \to 0$.]

References

Adams, A.M., Hudson, R.R., 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics 168 (3), 1699–1712.

Bellman, R., 1961. A Brief Introduction to Theta Functions. Holt, Rinehart and Winston.

Chen, H., Green, R., Pääbo, S., Slatkin, M., 2007. The joint allele-frequency spectrum in closely related species. Genetics 177, 387–398.

DeVore, R., Lorentz, G., 1993. Constructive Approximation. Springer.

Ewens, W., 1963. The diffusion equation and a pseudo-distribution in genetics. J. Roy. Stat. Soc. (B) 25, 405–412.

Garrigan, D., Hammer, M.F., 2006. Reconstructing human origins in the genomic era. Nat. Rev. Genet. 7 (9), 669–680.

Gordon, C., Webb, D., Wolpert, S., 1992. One cannnot hear the shape of a drum. Bull. Amer. Math. Soc. 27, 134–138.

Gradshteyn, I., Ryzhik, I., 1979. Table of Integrals, Series and Products. Academic Press.

Griffiths, R., 2003. The frequency spectrum of a mutation, and its age, in a general diffusion model. Theor. Popul. Biol. 64, 241–251.

Griffiths, R., Tavaré, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344 (1310), 403–410.

Kimura, M., 1955. Solution of a process of random genetic drift with a continuous model. PNAS 41, 144–150.

Kimura, M., Maruyama, T., 1975. Moments for sum of an arbitrary function of gene frequency along a stochastic path of gene frequency change. PNAS 72, 1602–1604.

Marth, G.T., Czabarka, E., Murvai, J., Sherry, S.T., 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166 (1), 351–372.

Nielsen, R., 2004. Population genetic analysis of ascertained SNP data. Hum. Genomics 1 (3), 218–224.

Patterson, N., 2005. How old is the most recent ancestor of two copies of an allele. Genetics 169, 1093–1104.

Przeworski, M., Hudson, R.R., Di Rienzo, A., 2000. Adjusting the focus on human variation. Trends Genet. 16 (7), 296–302.

Sawyer, S., 1977. On the past history of an allele now known to have frequency p. J. Appl. Probab. 14, 439–450.

Schaffner, S., Foo, C., Gabriel, S., Reich, D., Daly, M., Altshuler, D., 2005. Calibrating a coalescent simulation of human genome sequence variation. Genome Res. 15, 1576–1583.

Slatkin, M., Hudson, R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129 (2), 555–562.

Voight, B.F., Adams, A.M., Frisse, L.A., Qian, Y., Hudson, R.R., Di Rienzo, A., 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc. Natl. Acad. Sci. USA 102 (51), 18508–18513.

Wakeley, J., Nielsen, R., Liu-Cordero, S.N., Ardlie, K., 2001. The discovery of single-nucleotide polymorphisms–and inferences about human demographic history. Am. J. Hum. Genet. 69 (6), 1332–1347.

Watterson, G., 1976. Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. Theor. Popul. Biol. 10, 239–253.

Williamson, S.H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., Bustamante, C.D., 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. PNAS 102 (22), 7882–7887.