

An Estimate of the Average Number of Recessive Lethal Mutations Carried by Humans

Ziyue Gao,^{*1} Darrel Waggoner,^{†*} Matthew Stephens,^{†,§} Carole Ober,^{†,**} and Molly Przeworski^{††,**,1}

^{*}Committee on Genetics, Genomics, and Systems Biology, [†]Department of Human Genetics, [‡]Department of Pediatrics, [§]Department of Statistics, and ^{**}Department of Obstetrics and Gynecology, University of Chicago, Chicago, Illinois 60637, ^{††}Department of Biological Sciences, Columbia University, New York, New York 10027, and ^{**}Department of Systems Biology, Columbia University, New York, New York 10032

ABSTRACT The effects of inbreeding on human health depend critically on the number and severity of recessive, deleterious mutations carried by individuals. In humans, existing estimates of these quantities are based on comparisons between consanguineous and nonconsanguineous couples, an approach that confounds socioeconomic and genetic effects of inbreeding. To overcome this limitation, we focused on a founder population that practices a communal lifestyle, for which there is almost complete Mendelian disease ascertainment and a known pedigree. Focusing on recessive lethal diseases and simulating allele transmissions, we estimated that each haploid set of human autosomes carries on average 0.29 (95% credible interval [0.10, 0.84]) recessive alleles that lead to complete sterility or death by reproductive age when homozygous. Comparison to existing estimates in humans suggests that a substantial fraction of the total burden imposed by recessive deleterious variants is due to single mutations that lead to sterility or death between birth and reproductive age. In turn, comparison to estimates from other eukaryotes points to a surprising constancy of the average number of recessive lethal mutations across organisms with markedly different genome sizes.

KEYWORDS autosomal recessive disease; consanguinity; inbreeding; recessive lethal mutation; human

In diploid organisms such as humans, the efficacy of selection on a deleterious mutation depends both on the fitness of homozygotes and on the fitness of heterozygotes, which reflects dominance relationships among alleles. Since recently introduced mutations are mostly present in heterozygotes, they will be purged less effectively by selection when recessive and segregate at higher frequencies compared to dominant or semidominant alleles that cause a similar fitness reduction in homozygotes. For this reason, recessive alleles are expected to constitute a large fraction of strongly deleterious alleles segregating in diploid populations and in particular of Mendelian disease mutations in humans.

One context in which the effects of recessive mutations are unmasked is in the presence of inbreeding, which leads to an excess of homozygotes compared to Hardy–Weinberg expectation. Because closely related individuals may co-inherit

alleles from one or more common ancestors, the genomes of offspring of consanguineous couples are more likely to be identical by descent, revealing recessive, deleterious traits. If there are many recessive or partially recessive deleterious mutations (*i.e.*, mutations for which the fitness of the heterozygote is closer to that of the fitter homozygote) segregating in the population, inbred individuals will, on average, have lower fitness than outbred individuals. A reduction in mean fitness due to inbreeding (“inbreeding depression”) has been demonstrated repeatedly in experimental studies in multiple *Drosophila* species (*e.g.*, Greenberg and Crow 1960; reviewed in Simmons and Crow 1977), as well as under seminatural conditions in mice (Meagher *et al.* 2000).

Estimating the burden of recessive deleterious mutations in humans is therefore key to predicting adverse outcomes of consanguineous unions due to genetic factors (Morton *et al.* 1956; Bittles and Makov 1988; Bittles and Neel 1994; Bittles and Black 2010a). Two main methods have been developed to these ends: both aim to quantify the burden by comparing the health states of offspring of non-consanguineous and consanguineous matings. The first considers couples with variable degrees of relatedness and regresses the viabilities of their offspring on their inbreeding

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.114.173351

Manuscript received December 9, 2014; accepted for publication February 11, 2015; published Early Online February 18, 2015.

Available freely online through the author-supported open access option.

¹Corresponding authors: University of Chicago, 920 E. 58th St., CLSC 413, Chicago, IL 60637. E-mail: ziyuegao@uchicago.edu; Columbia University, 602A Fairchild Center, M.C. 2424, New York, NY 10027. E-mail: mp3284@columbia.edu

coefficients, F (Morton *et al.* 1956). When applied to humans, this method suffers from a number of limitations. For one, the estimate relies heavily on accurate assessments of degrees of relatedness, and yet the F values estimated from recent pedigrees do not capture inbreeding among more distant ancestors. This will bias the results if, as seems plausible, consanguineous marriages tend to occur in families with a tradition of close-kin unions (Hussain and Bittles 2000; Hamamy *et al.* 2011). Even if F is calculated based on deeper pedigrees, it represents only the expected proportion of the genome that is identical by descent, whereas the realized proportion could vary tremendously across individuals. In practice, this variation, combined with sampling variance, can lead to considerable uncertainty in the estimated effects of recessive alleles (Bittles and Makov 1985, 1988). Moreover, due to the restricted range of F and the small number of data points, the estimate of the combined effect of recessive deleterious mutations is highly sensitive to the choice of the regression model (Makov and Bittles 1986). Perhaps most importantly, consanguineous and non-consanguineous groups differ with respect to socioeconomic factors, in ways that influence the mortality and morbidity of the progeny (Schull and Neel 1965; Neel *et al.* 1970; Hussain and Bittles 1998, 2000). How estimates of genetic effects will be affected is unclear, as the strength of the correlation between socioeconomic status and inbreeding—and even the direction of the correlation—varies across societies (*e.g.*, Neel *et al.* 1970). Thus, this approach could either overestimate or underestimate the genetic effects of consanguinity on health outcomes.

To minimize these concerns, a second approach focuses specifically on the comparison between offspring of first-cousin marriages and of nonconsanguineous marriages in a large number of populations (Bittles and Makov 1988; Bittles and Neel 1994; Bittles and Black 2010b). Regression of the mortality of first-cousin progeny—for which the F value is 1/16—on that of nonconsanguineous progeny in the same population reveals a significant excess mortality in the former, which is then translated into the aggregate effect of recessive deleterious mutations. Even in this approach, however, genetic effects may be confounded by socioeconomic conditions that differ between consanguineous and nonconsanguineous groups *within* a population (Bittles and Neel 1994; Hussain and Bittles 1998).

Here, we introduce an approach that is not confounded by environmental effects, considering a founder population that practices a communal lifestyle, with a known pedigree and close to complete disease ascertainment over the past few generations (Hostetler 1974). Founder populations have contributed greatly to the identification of Mendelian disease mutations, because the founding bottleneck and subsequent inbreeding increase the chance of recent identity-by-descent and thus the incidence of a number of otherwise rare, recessive diseases (Boycott *et al.* 2008). With a known pedigree, we can estimate the probability that an autosomal recessive, deleterious founder mutation manifests itself (*i.e.*,

occurs as homozygous in at least one individual) in the past few generations, by simulating its transmission down the pedigree. From this estimate and the number of recessive diseases observed in the pedigree, we can obtain an estimate of the total number of deleterious mutations carried by the founders. Since the number of founders is known, this estimate translates into the average number of recessive lethal alleles in each haploid set of autosomes. An advantage of our approach is that, by directly utilizing the pedigree information, there is no need to calculate an inbreeding coefficient or to compare among groups that are potentially subject to different socioeconomic conditions. A difficulty, however, is that the transmission probability of a recessive deleterious allele depends on its selection coefficient in homozygotes (s), which is in general very hard to quantify. We therefore focused on autosomal, recessive lethal mutations ($s = 1$), defined as mutations that when homozygous lead to complete sterility or death between birth and reproductive age, in the absence of medical treatment. Since recessive lethal mutations are only a subset of all deleterious mutations, our estimate provides a lower bound on the burden of recessive deleterious mutations, as well as information on the tail of the distribution of fitness effects of deleterious mutations (*e.g.*, Yampolsky *et al.* 2005; Eyre-Walker *et al.* 2006; Boyko *et al.* 2008).

Materials and Methods

Simulations

To assess the probability of a recessive lethal mutation manifesting itself after 1950, we ran two sets of gene-dropping simulations.

Simulations on the minimum pedigree: We first considered a “minimum pedigree,” consisting of individuals ancestral to 1642 Hutterites for whom the genotypes at 14 autosomal recessive disease-causing genes were determined previously (Chong *et al.* 2012). When using this pedigree, we assumed no transmission distortion (Meyer *et al.* 2012) and complete reproductive compensation; *i.e.*, we assumed no correlation between the average number of surviving children and whether the parents are carriers of recessive lethals (Ober *et al.* 1999). If reproductive compensation is incomplete, our simulations are expected to lead to an underestimate of the fraction of recessive lethal alleles purged by selection since the founding and an overestimate of the probability of manifestation. In each replicate, we assigned a mutation to one founder and simulated the genotypes of other individuals generation by generation. For any individual who was known to have children, if both his (or her) parents were carriers of a recessive lethal, we assigned a heterozygous genotype with probability 2/3 and a homozygous beneficial genotype with probability 1/3. For all other situations, we simulated the offspring genotype based on the parental genotypes according to Mendelian inheritance rules. This

simulation scheme relies on the assumption of complete reproductive compensation, but is robust to missing data in the pedigree. After generating genotypes of all individuals in the pedigree, we examined the numbers of heterozygotes and homozygotes for that mutation among people born after 1950. The mutation was classified as “lost” if there was no heterozygote or homozygote in the cohort and as “manifested” if at least one homozygote was present in the cohort. Because individuals in the pedigree are related to each other, the incidences of a disease in the pedigree are not independent, so we considered the number of unique diseases instead of the number of affected individuals. We performed 320,000 gene-dropping simulations (5000 replicates for each founder). The mutation was lost before 1950 in 183,767 (57.4%) cases and manifested in 25,964 (8.11%) cases. We also considered all individuals born after 1940 as the cohort. Among the 320,000 simulations performed, the mutation was lost before 1940 in 182,221 (56.9%) cases and manifested in 26,328 (8.23%) cases. We focus on this simulation scheme in the main text.

Simulations on a larger pedigree: We also considered a larger pedigree, which consists of 15,236 individuals, all of whom can be traced back to 78 ancestors (who were not necessarily the “minimum ancestors” defined in Martin 1970). Individuals who fell into one of the following three groups were included in this pedigree: (1) before the separation of the three leuts, individuals who had descendants in the Schmiedeleut (S-leut); (2) all S-leut individuals who were born since the separation of the three leuts through 1980; and (3) S-leut individuals who were born between 1980 and 2013 and participated in our ongoing studies (Chong *et al.* 2012). When using this larger pedigree, we assumed no reproductive compensation and no transmission distortion (Meyer *et al.* 2012). Specifically, in each replicate, we assigned a mutation to one founder and then simulated the genotypes of all other individuals according to Mendelian inheritance rules. Because individuals homozygous for a recessive lethal mutation cannot reproduce, any individual who has offspring cannot be a homozygote recessive. To model this property of recessive lethals, we retained only replicates that were consistent with this condition, *i.e.*, replicates where all individuals with offspring were either heterozygous or homozygous for the beneficial allele. This simulation scheme would be exact on a complete pedigree, in which differences in family size are entirely attributable to recessive lethals and stochasticity. However, it is sensitive to incompleteness of the pedigree and other nonrandom factors that affect family size, in ways that are expected to lead to an overestimate of the fraction of recessive lethal alleles purged by selection since the founding and hence an underestimate of the probability of manifestation. We performed 78,000 gene-dropping simulations on the larger pedigree (1000 replicates for each founder). Among the 67,964 replicates retained, the mutation was lost before 1950 in 46,719 (68.7%) cases and manifested in 5563

(8.19%) cases. The proportion of recessive lethal mutations lost is higher than that of neutral variants (60.0%), suggesting that in addition to the dominant effect of genetic drift, selection could also have played a role in purging out recessive lethals in Hutterites. We also ran 39,000 replicates where we considered individuals born after 1940 as the cohort (500 replicates for each founder). Among the 34,001 replicates retained, the mutation was lost in 23,242 (68.4%) cases and manifested in 2787 (8.20%) cases.

For both scenarios, we also considered the situation where more than one copy of the same mutation was present in the founders. In each replicate, we randomly sampled two (or three) founders to be the carriers and simulated the genotypes of other individuals as described above. A total of 100,000 simulations were run with the minimum pedigree with two or three carriers, and the mutation was manifested in 16.7% and 25.5% of the cases, respectively. A total of 10,000 simulations were performed with the larger pedigree with two or three carriers, and the mutation was manifested in 14.8% and 20.6% of the cases retained, respectively. These findings indicate that the probability of manifestation is approximately proportional to the number of carriers among the founders, enabling us to estimate the total number of recessive lethal alleles (some of which could be copies of the same mutation) carried by the founders by dividing the number of distinct diseases by the probability of manifestation with one carrier.

Identification of recessive lethal diseases in Hutterites in the pedigree

Most Mendelian diseases reported in Hutterites were summarized in a review by Boycott *et al.* (2008), and the list of recessive diseases was further updated by Chong *et al.* (2012). To incorporate newly identified diseases, we searched PubMed for genetic diseases in Hutterites that were reported since 2012. We also searched for diseases with terms “recessive” and “Hutterites” in the Online Mendelian Inheritance in Man (OMIM) (an online catalog of human genetic disorders and underlying genes) and confirmed that all entries are included in our list.

We then classified a disease as “recessive lethal” if (i) it has 100% penetrance in homozygotes, (ii) the heterozygotes are asymptomatic (although they may have subtly decreased fitness), and (iii) the disease leads to prereproductive lethality (*e.g.*, restrictive dermopathy, cystic fibrosis in females) or complete sterility (*e.g.*, cystic fibrosis in males) in the absence of treatment. We included infertility due to biological reasons (*e.g.*, cystic fibrosis in males) or inability to reproduce if the phenotype of the condition leads to a reproductive fitness of zero due to social barriers (*e.g.*, nonsyndromic mental retardation, myopathy with movement disorder and intellectual disability).

To restrict the number of recessive lethal diseases to the pedigree under study, we required there to be affected individuals in Hutterites in South Dakota. This narrowed down the list to four diseases (cystic fibrosis, nonsyndromic

mental retardation, restrictive dermatopathy, and myopathy with movement disorder and intellectual disability). We excluded restrictive dermatopathy when considering the minimum pedigree, because the only reported patient in S-leut was not included in the minimum pedigree (although the parents were included and confirmed to be carriers by genotyping) (Chong *et al.* 2012; Loucks *et al.* 2012). For the other three diseases, genotype data of the 1642 extant individuals confirmed the presence of an individual(s) homozygous for the disease allele in the pedigree (Chong *et al.* 2012).

We note that two *CFTR* mutations have been identified in the Hutterites. Both alleles lead to severe phenotypes such that homozygous or compound heterozygous males are completely sterile and homozygous or compound heterozygous females cannot survive to reproductive age in absence of treatment. We therefore treat them as two copies of the same recessive lethal mutation. We further note that although the p.F508del mutation is common in Europeans, it is present on a single haplotype in Hutterites, suggesting that it was introduced into the population by only one founder (Chong *et al.* 2012). The p.M1101K mutation is rare in Europeans but was identified on two haplotypes in Hutterites (Zielenski *et al.* 1993). The two haplotypes differ at multiple sites on both sides of the mutation, indicating either that at least two recombination events have occurred or that the p.M1101K mutation was introduced by two founders (Chong *et al.* 2012). Therefore, it is likely that two or three carriers of these two *CFTR* mutations were present in the founders. Given that the probability of manifesting a mutation is approximately proportional to the number of carriers in the founders, we can treat it as introduced by only one founder.

Point estimation and credible intervals for the mean number of mutations per haploid human genome

We used a Bayesian approach to estimate the credible interval for mean number of recessive lethal alleles carried by each haploid set of human autosomes, R . Given that D recessive lethal diseases have been observed, the posterior probability of R is

$$f_{R|D}(r) = \frac{f_R(r)P(D|R=r)}{\int_{r=0}^{\infty} f_R(r)P(D|R=r)dr} \propto f_R(r)P(D|R=r), \quad (1)$$

where $f_{R|D}(r)$ is the conditional probability of observing D diseases, and $f_R(r)$ is the prior on R .

Let X_i be the number of unique recessive lethal mutations carried by the i th founder, among which Y_i mutations manifested themselves in the pedigree. We assume that each X_i is independently Poisson distributed:

$$X_i|R=r \sim \text{Poisson}(2r). \quad (2)$$

For simplicity, we assume all mutations carried by the founders are unique. If the transmission of each mutation is independent and each of the X_i mutations carried by the i th founder has a manifestation probability of p_i , then Y_i follows a binomial distribution:

$$Y_i|X_i \sim \text{Bin}(X_i, p_i). \quad (3)$$

Because of the thinning property of Poisson distribution, conditional on R , Y_i follows a Poisson distribution:

$$Y_i|R=r \sim \text{Poisson}(2rp_i).$$

Moreover, because each Y_i depends only on the corresponding X_i and, conditional on R , the X_i 's are independent of each other, the Y_i 's are independent Poisson random variables. Therefore, the total number of diseases observed, $D = Y_1 + Y_2 + \dots + Y_{N_f}$, is a Poisson random variable

$$D|R=r \sim \text{Poisson}(2rpN_f), \quad (4)$$

where $P = (p_1 + p_2 + \dots + p_{N_f})/N_f$ represents the probability of manifestation if each mutation was equally likely to be carried by each founder, which can be estimated from our simulations.

Combining (1) and (4), we can rewrite the posterior distribution of R as

$$\begin{aligned} f_{R|D}(r) &= \frac{f_R(r) \left((2rpN_f)^D / D! \right) e^{-2rpN_f}}{\int_{r=0}^{\infty} f_R(r) \left((2rpN_f)^D / D! \right) e^{-2rpN_f} dr} \\ &\propto f_R(r) \frac{(2rpN_f)^D}{D!} e^{-2rpN_f}. \end{aligned}$$

Because there is little definitive information about R , we use a uniform prior on $(0, \infty)$,

$$f_R(r) = 1,$$

as well as a uniform prior of the logarithm of R on $(0, \infty)$,

$$f_R(r) = \frac{1}{r}.$$

While both are improper priors, the resulting posterior distributions are proper gamma distributions:

$$f_{R|D}(r) \sim \text{Gamma}(D + 1, 2pN_f), \quad \text{if a uniform prior;}$$

or

$$f_{R|D}(r) \sim \text{Gamma}(D, 2pN_f), \quad \text{if a uniform prior on the logarithmic scale.}$$

Therefore, the posterior mode is $R = D/2pN_f$, if we assume a uniform prior [or $R = (D - 1)/2pN_f$, if a uniform prior on the logarithmic scale]. Given the value of p obtained from the simulations, the 95% credible interval (CI) of the posterior distribution was determined by using R (R Development Core Team 2013). For simulations with the minimum pedigree ($D = 3, P = 0.0811, N_f = 64$), the CI for R is [0.10, 0.84] when a uniform prior is assumed and

[0.060, 0.70] with a uniform prior on the logarithmic scale. Similar results are obtained from the larger pedigree ($D = 4$, $P = 0.0819$, $N_f = 78$): the CI for R is [0.13, 0.80] with a uniform prior and [0.085, 0.69] with a uniform prior on the logarithmic scale.

The target size for mutations that lead to recessive lethal disorders between birth and reproductive age

A genomic parameter of interest is the total number of functionally important sites that, if mutated, would give rise to recessive lethal alleles. We term this quantity the “target size” for recessive lethal disorders and infer it from our estimate of R .

The completely recessive case: Using a diffusion model and a low mutation rate approximation, Li and Nei (1972) derived the expectation for the total number of heterozygotes that carry a recessive deleterious mutation in a finite population [designated by $n_1(p)$],

$$n_1(p) = 4NN_eP\sqrt{\frac{\pi}{2N_e s}}, \quad \text{for } N_e s \geq 5$$

(equation 7b in Li and Nei 1972), where p is the initial frequency of the mutation, N and N_e the actual and effective population sizes, respectively, and s the selection coefficient against homozygotes (Li and Nei 1972).

The total number of heterozygotes affected by a single recessive lethal mutation, \bar{n}_1 , can be obtained by replacing p and s by $1/(2N)$ and 1 in the above equation:

$$\bar{n}_1 = \sqrt{2\pi N_e}.$$

For a gene with a mutation rate to recessive deleterious alleles of μ , the expected frequency of heterozygotes in a random generation is then

$$\bar{f}_1 = 2\mu\bar{n}_1$$

(equation 18 in Li and Nei 1972). For simplicity, we assume that there are m autosomal genes in the genome that each can lead to complete sterility or lethality between birth and reproductive age and that gene i has l_i sites at which mutations will give rise to recessive lethal alleles. We further assume that each site has the same per generation mutation rate, μ_{bp} , so the total mutation rate to recessive lethal alleles at gene i is

$$\mu_i = l_i\mu_{bp}.$$

The expected frequency of heterozygotes carrying a recessive lethal allele at this gene i is

$$\bar{f}_i = 2\mu_i\bar{n}_1 = 2\mu_i\sqrt{2\pi N_e},$$

and the expected frequency of recessive lethal alleles at this gene is approximately

$$\bar{q}_i = \frac{\bar{f}_i}{2} = \mu_i\sqrt{2\pi N_e}.$$

Therefore, the average number of recessive lethal mutations carried by a haploid genome is

$$\begin{aligned} \sum_{i=1}^m \bar{q}_i &= \sum_{i=1}^m \mu_i\sqrt{2\pi N_e} = \sum_{i=1}^m l_i\mu_{bp}\sqrt{2\pi N_e} \\ &= \mu_{bp}\sqrt{2\pi N_e} \sum_{i=1}^m l_i, \end{aligned}$$

where $\sum_{i=1}^m l_i$ is the target size for all autosomal recessive lethal mutations.

Assuming that the founders of S-leut Hutterites were drawn from a random-mating population at equilibrium, each of them should have carried twice that number of recessive lethals. Based on our estimate of $R = 0.29$, a mutation rate of $1.2 \times 10^{-8}/\text{bp}$ per generation (Campbell *et al.* 2012), and a diploid effective population size of 20,000, the target size is then estimated to be 6.8×10^4 bp.

In a population with larger N_e , the efficacy of purifying selection against recessive alleles increases (proportional to $N_e^{1/2}$), but the increase in the mutational input (proportional to N_e) is greater than the increase in the efficacy of selection (Simons *et al.* 2014). Therefore, at equilibrium, an individual in a population with larger N_e will have a greater expected number of recessive lethal mutations. The recent population growth experienced by humans represents a transition from small N_e to large N_e , which will therefore lead to an increase in the average number of recessive lethal mutations per individual. As a result, the estimated target size after recent growth will be smaller than that estimated from the long-term N_e .

The partially recessive case: If a deleterious mutation leads to complete lethality (or sterility) in homozygotes ($s = 1$) and a decrease of hs in fitness of heterozygotes, selection against the deleterious allele would mainly come from the death of heterozygotes, because the death of homozygotes is a rare event (when the allele frequency is low). As in the completely recessive case, we assume that there are m autosomal genes in the genome that can mutate to deleterious alleles with these properties and that gene i has l_i such sites. Under these assumptions, the expected total number of heterozygotes affected by a mutation in a finite population at gene i can be approximated as

$$\bar{n}_1 = \frac{1}{hs} = \frac{1}{h}, \quad \text{for } 4N_e sh \geq 1$$

(equation 6b in Li and Nei 1972). So the average frequency of those partially recessive lethal mutations is

$$\bar{q}_i = \mu_i\bar{n}_1 = \frac{\mu_i}{h},$$

which is independent of the effective population size and the same as obtained assuming mutation–selection balance (Haldane 1935).

Therefore, the total number of such partially recessive lethal mutations carried by a random haploid genome is

$$\sum_{i=1}^m \bar{q}_i = \sum_{i=1}^m \frac{\mu_i}{h} = \sum_{i=1}^m \frac{l_i \mu_{bp}}{h} = \frac{\mu_{bp}}{h} \sum_{i=1}^m l_i,$$

where $\sum_{i=1}^m l_i$ is the target size for all mutations that lead to lethality in homozygotes and a fitness decrease of hs in heterozygotes. Plugging in $s = 1$, $h = 0.01$ and taking a mutation rate of $1.2 \times 10^{-8}/bp$ per generation (Campbell *et al.* 2012), we estimate that the target size of such nearly recessive mutations is 2.4×10^5 bp.

Results

We focused on the Hutterites, a group ideally suited for the new method we propose. The Hutterian Brethren is an isolated founder population, which originated in the Tyrolean Alps in the 1520s and eventually settled in North America on three communal farms in the 1870s after a series of migrations throughout Europe. The three colonies thrived and shortly thereafter gave rise to three major subdivisions, referred to as the Schmiedeleut (S-leut), Lehrerleut (L-leut) and Dariusleut (D-leut), with most marriages since 1910 taking place among individuals within the same leut. Each leut practices a communal lifestyle, with no private ownership and hence no socioeconomic differences among members (Hostetler 1974).

The Hutterites have kept extensive genealogical records, from which highly reliable pedigrees have been reconstructed (Bleibtreu 1964; Mange 1964; Steinberg *et al.* 1967). Moreover, researchers and the Hutterites community have built a close partnership, greatly facilitating the diagnosis and identification of genetic disorders. Incidences of genetic disorders in Hutterites have been comprehensively documented since the late 1950s, with >40 Mendelian disorders, of which 35 are autosomal recessive, described in the literature (Boycott *et al.* 2008, 2010; Çalışkan *et al.* 2011; Huang *et al.* 2011; Bögershausen *et al.* 2013; Gerull *et al.* 2013; Wiltshire *et al.* 2013). The pedigree information and nearly complete ascertainment of genetic disorders over the past ~60 years make it possible to reliably infer the number of recessive lethal mutations in the founders.

Specifically, our analysis was based on a 13-generation pedigree that relates 1642 extant S-leut Hutterites in South Dakota and their ancestors (3657 individuals in total), all of whom can be traced back to 64 founders who lived in the early 18th to early 19th centuries in Europe (Chong *et al.* 2012). We ran gene-dropping simulations (similar to those in Manatrion *et al.* 2009 and Chong *et al.* 2012) on the pedigree to assess the probability of an autosomal recessive lethal mutation manifesting itself, conditional on being carried by a founder (see *Materials and Methods* for simulation procedures). We defined “manifestation” as the presence of at least one individual in the pedigree who was born after 1950—the period after which disease ascertainment is close

to complete—and inherited two copies of a recessive lethal variant introduced by the founder.

We found that on average 57% of unique recessive lethal mutations carried by the founders have been lost before 1950 (*i.e.*, none of the individuals in the pedigree born after 1950 carries the mutation). This proportion is almost the same as for neutral variants with the same initial frequency in the founders, suggesting that the loss of variants is primarily due to the severe genetic drift after the founder event (confirming results found in Chong *et al.* 2012). Among the recessive lethal mutations that survive until 1950, we expect that 19% will have been exposed in homozygote(s) and thus, in expectation, 8.1% of all recessive lethal mutations carried by the founders will have manifested themselves to date. The probability is almost the same if we consider individuals born after 1940 as the cohort, to allow for a delay in diagnosis for diseases with an onset in adolescence (see *Materials and Methods*).

The above simulation scheme implicitly assumes complete reproductive compensation (see *Materials and Methods*), which might not be appropriate for all recessive lethal diseases. To address this concern, we ran a second set of simulations on a larger pedigree of 15,235 Hutterites, who can be traced back to 78 ancestors (see *Materials and Methods* for details about this pedigree). The second simulation scheme assumes no reproductive compensation and would be exact on a complete pedigree in which differences in family size are entirely attributable to recessive lethals and stochasticity, but it is sensitive to the incompleteness of the pedigree (especially missing data in the early generations). Nonetheless, the results are similar: on average, 69% of recessive lethal mutations carried by the founders will have been lost before 1950 and 8.2% will have manifested themselves in individuals born after 1950.

Next, we considered all known autosomal recessive lethal diseases observed in the S-leut Hutterites included in the pedigree. To this end, we compiled a list of all known autosomal recessive disorders (11 in total) reported in S-leut Hutterites in the United States (see *Materials and Methods* for details). We classified each of the identified diseases as “lethal” if, untreated, it leads to death between birth and reproductive age or precludes reproduction for affected individuals. We found four such recessive lethal diseases: cystic fibrosis, nonsyndromic mental retardation, a severe form of myopathy, and restrictive dermopathy (Table 1). The underlying mutations are known for all four diseases, and genotyping data of the 1642 extant individuals confirmed the presence of homozygote(s) of the disease-causing mutations for the first three diseases (Chong *et al.* 2012). Restrictive dermopathy was excluded from our list when we used the 13-generation pedigree, because the only known case among the South Dakota Hutterites was not included in the 3657 individuals.

From the number of recessive lethal diseases (three) and the probability that a recessive lethal allele manifested itself since 1950 (0.081), we estimated that the total number of

Table 1 Three autosomal recessive lethal diseases and corresponding mutations observed in the smaller pedigree

Name of disease	OMIM phenotype no.	Gene	Mutation	Carrier frequency in Europeans	Carrier frequency in S-leut Hutterites (%)	Source
Cystic fibrosis ^a	219700	<i>CFTR</i>	p.F508del ^a p.M1101K ^a	2.2–3.9% Unknown (one case reported)	2.2 ^b 7.3 ^b	Fujiwara <i>et al.</i> (1989); Zielenski <i>et al.</i> (1993)
Nonsyndromic mental retardation	614020	<i>TECR</i>	p.P182L	Found only in Hutterites to date	6.9 ^b	Çalışkan <i>et al.</i> (2011)
Myopathy with movement disorder and intellectual disability	—	<i>TRAPPC11</i>	c.1287+5G > A	Found only in Hutterites to date	7	Bögershausen <i>et al.</i> (2013)

^a See *Materials and Methods* for treatment of the two mutations in *CFTR*.

^b The allele frequencies are reported in Chong *et al.* (2012).

autosomal, recessive lethal mutations carried by the 64 founders is $3/0.081 = 37$ or an average of 0.29 recessive lethal alleles in each haploid human genome (Figure 1A). To assess the uncertainty in this estimate, we estimated the posterior distribution of the mean number of mutations per haploid human genome conditional on observing exactly three diseases since 1950 (*Materials and Methods* and Figure 1B). If a uniform prior distribution is used, the posterior distribution has a mode of 0.29 and a 95% CI of [0.10, 0.84]. We also considered a uniform prior on the logarithmic scale to allow for uncertainty in the order of magnitude, and a similar 95% CI is obtained (*i.e.*, [0.060, 0.70] mutations per haploid genome). The point estimate and 95% CI were similar, when we used simulation results from the larger pedigree (see *Materials and Methods*).

Simulations further indicate that only a small fraction of the surviving recessive lethal mutations have been seen in homozygotes, so there are more hidden, recessive lethal mutations that are segregating among extant individuals in the pedigree. In fact, carrier screening has identified heterozygotes for three more recessive lethal mutations in the S-leut Hutterites in South Dakota, which have manifested themselves in Hutterites outside the pedigree under study (Table 2) (Chong *et al.* 2012). Based on our simulation results, we expect quite a few more recessive lethal mutations in addition to these cases, most of which remain unknown.

In generalizing from the results for the Hutterites to other human populations, one concern might be that their demographic history prior to the founder event in the 18th–19th centuries was atypical in ways that influence the number of recessive lethals carried by the founders. While transient demographic changes can have a marked impact on patterns of genetic variation, they are not expected to have a substantial effect on the average number of recessive lethal alleles carried by an individual, because their equilibrium frequency is reestablished on a relatively short timescale (Balick *et al.* 2014). For instance, after a bottleneck, this quantity of interest returns to the equilibrium value within $\sqrt{4N_0}$ generations (where N_0 is the original population size before the bottleneck); for this reason, this quantity is expected to be very similar for modern Africans

and Europeans despite the out-of-Africa bottleneck. A similar concern might be that a long period of endogamy could have purged recessive deleterious alleles from the population that led to the Hutterites (Keller and Waller 2002). There is no evidence that such a demographic scenario occurred, but if it did, it is again unlikely to have had much of an effect: over the ~15 generations between the origin of the Hutterites in the 1520s and the founding event, even relatively high levels of human inbreeding ($F = 0.03$) should decrease the mean allele frequency of recessive lethals only by ~30% (Overall *et al.* 2002). Moreover, such a decrease would be lessened or nullified by reproductive compensation (Overall *et al.* 2002), as might occur in the Hutterites (Ober *et al.* 1999). These considerations suggest that estimates based on the Hutterites should be broadly applicable and would, if anything, be slightly lower than the mean number of recessive lethals carried by larger, outbred populations.

In that regard, we note that our estimate of the average number of recessive lethal mutations per haploid genome is lower than the previous estimates of the total number of “lethal equivalents” per haploid genome (0.56–0.7 in Bittles and Neel 1994; Bittles and Black 2010b). A lethal equivalent is defined as a locus or a set of loci that, when in the homozygous state, would cause on average one death, *e.g.*, one lethal mutation or two mutations each with 50% probability of causing death (Morton *et al.* 1956). In other words, the total number of lethal equivalents in a haploid genome can be thought of as the sum of the deleterious effects of *all* recessive mutations carried by an individual. Comparison to estimates of this quantity suggests that, as expected, recessive lethal mutations are only a subset of the recessive mutational burden. Interestingly, however, the difference between our point estimate and previous estimates is only about twofold; even if we consider the lower bound of our credible interval on the mean number of recessive lethals, it is still about one-sixth of the total number of lethal equivalents. Thus, insofar as previous estimates are reliable, it appears that a substantial portion of the total burden of recessive mutations carried by humans is attributable to single mutations that, when homozygous, lead to sterility or death between birth and reproductive age.

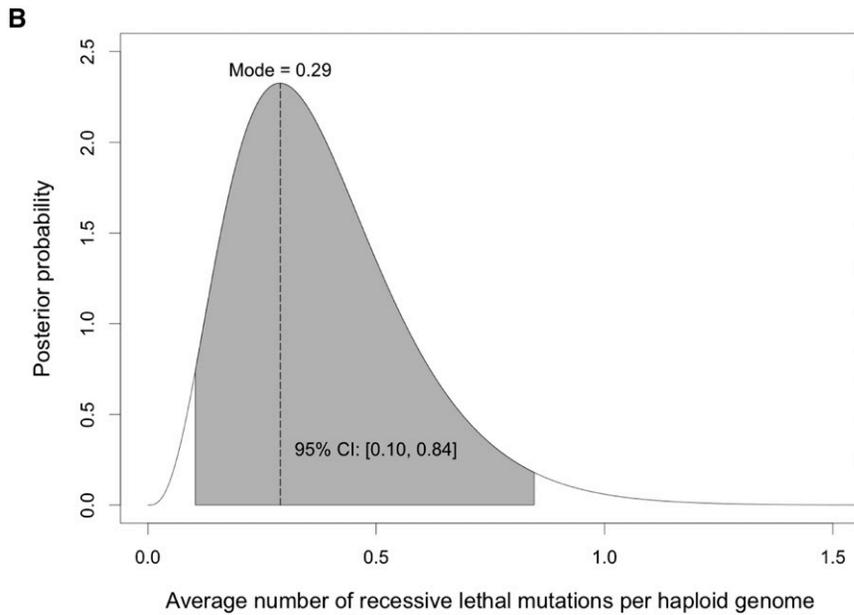
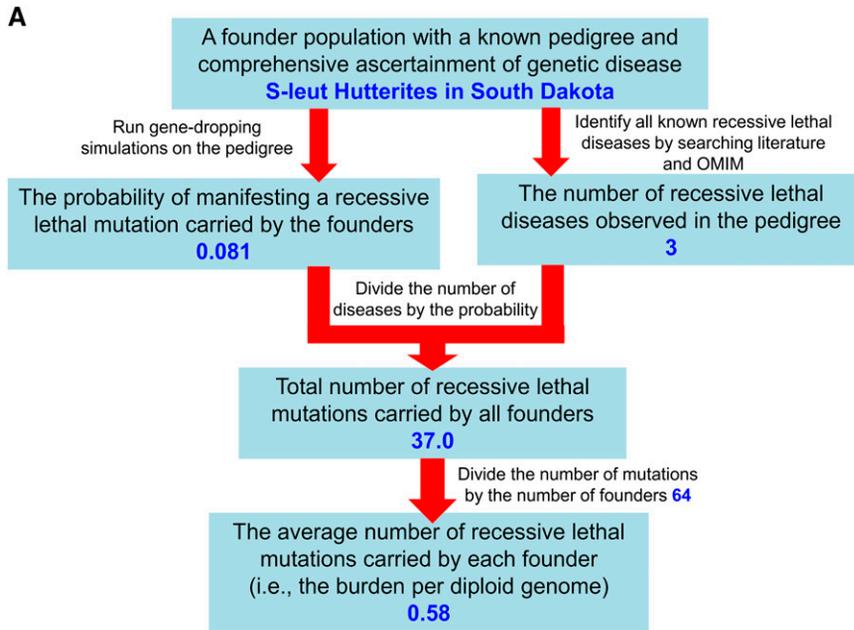


Figure 1 The analysis pipeline for estimating the average number of recessive lethal mutations carried by humans. (A) A schematic diagram of the approach. The analysis procedures are described in black. Specific values and estimates for the Hutterites are provided in blue, including the point estimate of the mean number of mutations carried by a founder. (B) The posterior distribution of the mean number of recessive lethal mutations carried by each haploid human genome, given the probability of manifestation and the number of diseases observed.

Discussion

Our approach indicates that on average, each Hutterite founder carried ~0.58 autosomal, recessive lethal mutations that lead to death between birth and reproductive age or to complete sterility. This is likely a slight underestimate for other human populations, if we take into account the effects of the unique demographic history of the Hutterites. Nonetheless, this estimate is unaffected by socioeconomic factors. Moreover, incomplete ascertainment of diseases is unlikely to be a major concern, because most severe genetic disorders that occurred in Hutterites after the 1950s are expected to have been documented (Boycott *et al.* 2008), so we expect at most a slight underestimate of their number. In addition, while we ignore interference and linkage between recessive lethal alleles when

simulating the transmission of those mutations, this should not influence the mean proportion of mutations that survive or manifest themselves, so will not bias our estimate of the probability of manifestation. We also ignore the possibility that *de novo* mutations that arose since the founding may contribute to the diseases considered. This assumption is justified because we expect at most a few recessive lethal mutations to have arisen in the pedigree in the 13 generations (see discussion on the target size of recessive lethals below) and their probability of manifestation is even lower than that of founder mutations. Thus, overall, we expect the estimate to be relatively unbiased and, if anything, a slight underestimate of the mean value of other populations.

If we take our point estimate at face value, it suggests that the risk of autosomal recessive lethal disorders that

Table 2 Three autosomal recessive lethal mutations found only in heterozygotes in the pedigree of S-leut Hutterites in South Dakota

Name of disease	OMIM phenotype no.	Gene	Mutation	Carrier frequency in Europeans	Carrier frequency in S-leut Hutterites	Source
Bardet–Biedl syndrome	209900	<i>BBS2</i>	c.472-2A > G	Found only in Hutterites to date	2.8 ^a	Innes <i>et al.</i> (2010)
Dilated cardiomyopathy with ataxia syndrome	610198	<i>DNAJC19</i>	IVS3-1G > C	Found only in Hutterites to date	2.8 ^a	Davey <i>et al.</i> (2006)
Joubert syndrome/Meckel syndrome	614424	<i>TMEM237</i>	p.R18X	Found only in Hutterites to date	8.0 ^a	Schurig <i>et al.</i> (1980); Huang <i>et al.</i> (2011)

^a The allele frequencies are reported in Chong *et al.* (2012).

manifest after birth should be increased by $0.29/16 = 1.8\%$ in offspring of first-cousin couples (assuming no difference in environmental factors). This prediction agrees well with the estimated 3.5–4.4% increased risk for prereproductive mortality and 1.7–2.8% increased incidence of congenital anomalies in children of first cousins above the general population risk (Hamamy *et al.* 2011).

Beyond the Hutterites, this approach can be applied to other isolated founder populations with limited immigration, for which there is reliable genealogical information since the founding and close to complete disease phenotyping in the relatively recent past, such as the Amish and the inhabitants of Norfolk Island (Macgregor *et al.* 2010; Hinckley *et al.* 2013).

In interpreting our estimates, an important consideration is that they are limited to lethal diseases that manifest themselves after birth. This issue is common to most studies that estimate the mutational burden in humans, because of the limited availability and reliability of data on prenatal loss. Studies that considered data on the frequency of miscarriages (*i.e.*, a gestation age of ≥ 28 weeks) reported no or little effect of consanguinity on prenatal losses, while detecting clear-cut effects on postnatal mortality (Schull *et al.* 1970; Bittles and Makov 1988; Bittles and Black 2010b). This negative finding cannot be taken as strong evidence for the absence of embryonic recessive lethals in humans, as most losses due to embryonic lethals may occur during earlier stages of pregnancy. Even if the data on early pregnancy loss were available, the high rate of spontaneous pregnancy failure due to other causes (Leridon 1977) may obscure the difference between consanguineous and nonconsanguineous groups due to embryonic recessive lethals. In contrast to how little is known in humans, extensive mutation screens in mice reveal a high proportion (40–50%) of autosomal knockout mutations that cause deaths in prenatal stages when homozygous (Mitchell *et al.* 2001; White *et al.* 2013). If the proportion of embryonic lethals is similar for spontaneous mutations in humans, each human individual carries approximately one to two recessive lethal mutations that act across ontogenesis.

Genomic target size for recessive lethal mutations

Our results provide insight into the total number of autosomal sites in the human genome that, if mutated, give rise to recessive (or nearly recessive) lethal alleles. Those

sites are of particular interest, because they are of critical functional importance; on the other hand, mutations at those sites are haplosufficient, in that one functional copy of the gene is enough to maintain fitness. Assuming a random-mating, diploid population with constant effective population size of 20,000 (as a proxy for the population from which Hutterite founders derived), a mutation rate of $1.2 \times 10^{-8}/\text{bp}$ per generation, and an estimate of 0.29 recessive lethal mutations per haploid set of autosomes, we predict that there should be $\sim 68,000$ autosomal base pairs at which mutations lead to recessive lethal disorders on or after birth (see *Materials and Methods*; using theory from Li and Nei 1972). Based on this estimate of the target size, we do not expect *de novo* recessive lethals that arose since the founding to manifest themselves as diseases since the 1950s. Considering a model of population growth that leads the current effective population size to be $>20,000$, the expected frequency of recessive lethal mutations in the population should be higher (see *Materials and Methods*), and the estimated size of target size smaller. While these estimates should not be taken too literally, as many recessive disease mutations are not point mutations (*e.g.*, Boone *et al.* 2013), they provide a sense of the number of sites at which recessive lethal disease mutations may be discovered.

Moreover, this estimate of target size provides complementary information to population genetic approaches that aim to estimate the distribution of fitness effects of mutations from polymorphism and divergence and mostly learn about weaker selection coefficients (Eyre-Walker and Keightley 2007). These methods find that 25–40% of all amino acid changes in humans are strongly deleterious (*i.e.*, have $s > 1\%$ in a genic selection model) (Yampolsky *et al.* 2005; Eyre-Walker *et al.* 2006; Boyko *et al.* 2008). Combining these estimates with our estimated target size would then suggest that 0.5–0.8% of strongly deleterious mutations are recessive lethals that are fatal between birth and reproductive age.

An important caveat is that recessive disease-causing mutations may not be completely recessive, in that carriers of one copy may also have a slight decrease in their fitness that is too subtle to be detectable in clinical diagnosis. If so, the mutations will segregate in the population at much lower frequencies due to selection against heterozygotes, and the target size could be larger. For instance, if the

heterozygous effect is a 1% decrease in fitness, the corresponding target size would become $\sim 240,000$ bp (see *Materials and Methods*).

Comparison to other species

Intriguingly, our estimate of the average number of recessive lethal mutations per individual is in good agreement with what has been determined experimentally in a number of other diploid animal species. Most studies were conducted in *Drosophila melanogaster*, where individuals from natural or laboratory populations were made homozygous to measure the effects on viability. The results are relatively consistent, with on average 24.7% of the second chromosomes and 40.7% of the third chromosomes in the population carrying at least one recessive lethal (or nearly lethal) mutation (Simmons and Crow 1977). Assuming the number of such mutations is Poisson distributed, this implies that each *D. melanogaster* harbors on average ~ 1.6 autosomal recessive lethal mutations. Similar numbers were obtained in other *Drosophila* species (e.g., Dobzhansky *et al.* 1963; Malogolowkin-Cohen *et al.* 1964), as well as by sib crosses in *Lucania goodie* (1.87 lethal mutations per individual) and *Danio rerio* (1.43 mutations) (McCune *et al.* 2002).

Our estimate in humans (of approximately one to two recessive lethals that act across all developmental stages per diploid) is again quite similar, reviving the question of why such distantly related organisms carry similar numbers of recessive lethal mutations per genome, despite their highly variable genome sizes (McCune *et al.* 2002; Halligan and Keightley 2003). Under a model of mutation–selection balance, the equilibrium frequency of a recessive lethal allele in an outbred population depends solely on the mutation rate (Gillespie 2004). Given that the mutation rates per base pair per generation are thought to be similar across vertebrates and *Drosophila* (Ségurel *et al.* 2014), organisms with larger genome sizes are expected to carry more recessive lethals. This seemingly surprising constancy is reminiscent of the C-value paradox—the observation that the genome size of eukaryotes appears to reflect neither the organismal complexity nor the gene number (Thomas 1971). Although both observations suggest that the genome size is a poor predictor of the functional content of a genome, our finding highlights another aspect: the number of sites of crucial functional importance may be unexpectedly constant across taxa, despite the differences in the number of genes and the length of the coding regions (Alexander *et al.* 2010; Howe *et al.* 2013). Another potential contributing factor to the finding in humans could be the mating patterns. Over the past centuries, consanguineous marriages were a common practice, and the custom remains prevalent in many countries (Bittles and Black 2010a; Hamamy *et al.* 2011). In the long term, depending on the degree of reproductive compensation (Overall *et al.* 2002), inbreeding can facilitate the purging of recessive deleterious mutations from the population (Keller and Waller 2002), leading to a lower equilibrium mean

frequency of recessive lethal mutations in humans compared with more outbred organisms.

Acknowledgments

We thank the Hutterites for their ongoing support and participation in our studies. We thank Jessica Chong for helpful comments regarding the pedigree data and simulations; Mark Abney for discussing the simulation procedures; Kym Boycott and Micheil Innes for their help in clarifying which recessive lethal diseases were present in S-leut Hutterites; and Graham Coop, Dick Hudson, Marty Kreitman, Charles Langley, Noah Rosenberg, and Guy Sella for helpful discussions. This study was supported in part by grants R01 HD21244 and R01 HL085197. Z.G. was supported in part by the William Rainey Harper Fellowship from the University of Chicago.

Literature Cited

- Alexander, R. P., G. Fang, J. Rozowsky, M. Snyder, and M. B. Gerstein, 2010 Annotating non-coding regions of the genome. *Nat. Rev. Genet.* 11: 559–571.
- Balick, D. J., R. Do, D. Reich and S. Sunyeav, 2014 Response to a population bottleneck can be used to infer recessive selection. arXiv: 1312.3028.
- Bittles, A. H., and M. L. Black, 2010a Consanguinity, human evolution, and complex diseases. *Proc. Natl. Acad. Sci. USA* 107: 1779–1786.
- Bittles, A. H., and M. L. Black, 2010b The impact of consanguinity on neonatal and infant health. *Early Hum. Dev.* 86: 737–741.
- Bittles, A. H., and U. E. Makov, 1985 Linear regressions in the calculation of lethal gene equivalents in man. *Ann. Hum. Biol.* 12: 287–289.
- Bittles, A. H., and E. Makov, 1988 Inbreeding in human populations: an assessment of the costs, pp. 153–167 in *Human Mating Patterns*, edited by C. G. N. Mascie-Taylor, and A. J. Boyce. Cambridge University Press, Cambridge, UK/New York.
- Bittles, A. H., and J. V. Neel, 1994 The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* 8: 117–121.
- Bleibtreu, H., 1964 *Marriage and Residence Patterns in a Genetic Isolate*. PhD Thesis, Anthropology Department, Harvard University, Cambridge, MA.
- Bögershausen, N., N. Shahrzad, J. X. Chong, J. C. von Kleist-Retzow, D. Stanga *et al.*, 2013 Recessive TRAPPC11 mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. *Am. J. Hum. Genet.* 93: 181–190.
- Boone, P. M., I. M. Campbell, B. C. Baggett, Z. T. Soens, M. M. Rao *et al.*, 2013 Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. *Genome Res.* 23: 1383–1394.
- Boycott, K. M., J. S. Parboosingh, B. N. Chodirker, R. B. Lowry, D. R. McLeod *et al.*, 2008 Clinical genetics and the Hutterite population: a review of Mendelian disorders. *Am. J. Med. Genet. A.* 146A: 1088–1098.
- Boycott, K. M., C. Beaulieu, E. G. Puffenberger, D. R. McLeod, J. S. Parboosingh *et al.*, 2010 A novel autosomal recessive malformation syndrome associated with developmental delay and distinctive facies maps to 16p11 in the Hutterite population. *Am. J. Med. Genet. A.* 152A: 1349–1356.

- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4: e1000083.
- Çalışkan, M., J. X. Chong, L. Uricchio, R. Anderson, P. Chen *et al.*, 2011 Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the *TECR* gene on chromosome 19p13. *Hum. Mol. Genet.* 20: 1285–1289.
- Campbell, C. D., J. X. Chong, M. Malig, A. Ko, B. L. Dumont *et al.*, 2012 Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* 44: 1277–1281.
- Chong, J. X., R. Ouwenga, R. L. Anderson, D. J. Waggoner, and C. Ober, 2012 A population-based study of autosomal-recessive disease-causing mutations in a founder population. *Am. J. Hum. Genet.* 91: 608–620.
- Davey, K. M., J. S. Parboosingh, D. R. McLeod, A. Chan, R. Casey *et al.*, 2006 Mutation of *DNAJC19*, a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition. *J. Med. Genet.* 43: 385–393.
- Dobzhansky, T., B. Spassky, and T. Tidwell, 1963 Genetics of natural populations. 32. Inbreeding and the mutational and balanced genetic loads in natural populations of *Drosophila pseudoobscura*. *Genetics* 48: 361–373.
- Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8: 610–618.
- Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900.
- Fujiwara, T. M., K. Morgan, R. H. Schwartz, R. A. Doherty, S. R. Miller *et al.*, 1989 Genealogical analysis of cystic fibrosis families and chromosome 7q RFLP haplotypes in the Hutterite Brethren. *Am. J. Hum. Genet.* 44: 327–337.
- Gerull, B., F. Kirchner, J. X. Chong, J. Tagoe, K. Chandrasekharan *et al.*, 2013 Homozygous founder mutation in *desmocollin-2* (*DSC2*) causes arrhythmogenic cardiomyopathy in the Hutterite population. *Circ. Cardiovasc. Genet.* 6: 327–336.
- Gillespie, J. H., 2004 *Population Genetics: A Concise Guide*. Johns Hopkins University Press, Baltimore.
- Greenberg, R., and J. F. Crow, 1960 A comparison of the effect of lethal and detrimental chromosomes from *Drosophila* populations. *Genetics* 45: 1153–1168.
- Haldane, J. B., 1935 The rate of spontaneous mutation of a human gene. *J. Genet.* 31: 317–326.
- Halligan, D. L., and P. D. Keightley, 2003 How many lethal alleles? *Trends Genet.* 19: 57–59.
- Hamamy, H., S. E. Antonarakis, L. L. Cavalli-Sforza, S. Temtamy, G. Romeo *et al.*, 2011 Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. *Genet. Med.* 13: 841–847.
- Hinckley, J. D., D. Abbott, T. L. Burns, M. Heiman, A. D. Shapiro *et al.*, 2013 Quantitative trait locus linkage analysis in a large Amish pedigree identifies novel candidate loci for erythrocyte traits. *Mol. Genet. Genomic Med.* 1: 131–141.
- Hostetler, J. A., 1974 *Hutterite Society*. Johns Hopkins University Press, Baltimore.
- Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot *et al.*, 2013 The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496: 498–503.
- Huang, L., K. Szymanska, V. L. Jensen, A. R. Janecke, A. M. Innes *et al.*, 2011 *TMEM237* is mutated in individuals with a Joubert syndrome related disorder and expands the role of the *TMEM* family at the ciliary transition zone. *Am. J. Hum. Genet.* 89: 713–730.
- Hussain, R., and A. H. Bittles, 1998 The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *J. Biosoc. Sci.* 30: 261–275.
- Hussain, R., and A. H. Bittles, 2000 Sociodemographic correlates of consanguineous marriage in the Muslim population of India. *J. Biosoc. Sci.* 32: 433–442.
- Innes, A. M., K. M. Boycott, E. G. Puffenberger, D. Redl, I. M. MacDonald *et al.*, 2010 A founder mutation in *BBS2* is responsible for Bardet-Biedl syndrome in the Hutterite population: utility of SNP arrays in genetically heterogeneous disorders. *Clin. Genet.* 78: 424–431.
- Keller, L., and D. Waller, 2002 Inbreeding effects in wild populations. *Trends Ecol. Evol.* 17: 230–241.
- Leridon, H., 1977 *Human Fertility: The Basic Components*. University of Chicago Press, Chicago.
- Li, W. H., and M. Nei, 1972 Total number of individuals affected by a single deleterious mutation in a finite population. *Am. J. Hum. Genet.* 24: 667–679.
- Loucks, C., J. S. Parboosingh, J. X. Chong, C. Ober, V. M. Siu *et al.*, 2012 A shared founder mutation underlies restrictive dermopathy in Old Colony (Dutch-German) Mennonite and Hutterite patients in North America. *Am. J. Med. Genet. A.* 158A: 1229–1232.
- Macgregor, S., C. Bellis, R. A. Lea, H. Cox, T. Dyer *et al.*, 2010 Legacy of mutiny on the Bounty: founder effect and admixture on Norfolk Island. *Eur. J. Hum. Genet.* 18: 67–72.
- Makov, E., and A. H. Bittles, 1986 On the choice of mathematical models for the estimation of lethal gene equivalents in man. *Heredity* 57(Pt 3): 377–380.
- Malogolowkin-Cohen, C., H. Levene, N. P. Dobzhansky, and A. S. Simmons, 1964 Inbreeding and the mutational and balanced loads in natural populations of *Drosophila willistoni*. *Genetics* 50: 1299–1311.
- Manatrion, S., C. Egger-Danner, and R. Baumung, 2009 Estimating lethal allele frequencies in complex pedigrees via gene dropping approach using the example of Brown Swiss cattle. *Arch. Anim. Breed.* 52: 230–242.
- Mange, A. P., 1964 Growth and inbreeding of a human isolate. *Hum. Biol.* 36: 104–133.
- Martin, A. O., 1970 The founder effect in a human isolate: evolutionary implications. *Am. J. Phys. Anthropol.* 32: 351–367.
- McCune, A. R., R. C. Fuller, A. A. Aquilina, R. M. Dawley, J. M. Fadoo *et al.*, 2002 A low genomic number of recessive lethals in natural populations of bluefin killifish and zebrafish. *Science* 296: 2398–2401.
- Meagher, S., D. J. Penn, and W. K. Potts, 2000 Male-male competition magnifies inbreeding depression in wild house mice. *Proc. Natl. Acad. Sci. USA* 97: 3324–3329.
- Meyer, W. K., B. Arbeithuber, C. Ober, T. Ebner, I. Tiemann-Boege *et al.*, 2012 Evaluating the evidence for transmission distortion in human pedigrees. *Genetics* 191: 215–232.
- Mitchell, K. J., K. I. Pinson, O. G. Kelly, J. Brennan, J. Zupicich *et al.*, 2001 Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat. Genet.* 28: 241–249.
- Morton, N. E., J. F. Crow, and H. J. Muller, 1956 An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. USA* 42: 855–863.
- Neel, J. V., W. J. Schull, M. Yamamoto, S. Uchida, T. Yanase *et al.*, 1970 The effects of parental consanguinity and inbreeding in Hirado, Japan. II. Physical development, tapping rate, blood pressure, intelligence quotient, and school performance. *Am. J. Hum. Genet.* 22: 263–286.
- Ober, C., T. Hyslop, and W. W. Hauck, 1999 Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am. J. Hum. Genet.* 64: 225–231.
- Overall, A. D., M. Ahmad, and R. A. Nichols, 2002 The effect of reproductive compensation on recessive disorders within consanguineous human populations. *Heredity* 88: 474–479.

- R Development Core Team, 2013 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Schull, W. J., and J. V. Neel, 1965 *The Effects of Inbreeding on Japanese Children*. Harper & Row, New York.
- Schull, W. J., H. Nagano, M. Yamamoto, and I. Komatsu, 1970 The effects of parental consanguinity and inbreeding in Hirado, Japan. I. Stillbirths and prereproductive mortality. *Am. J. Hum. Genet.* 22: 239–262.
- Schurig, V., P. Bowen, F. Harley, and D. Schiff, 1980 The Meckel syndrome in the Hutterites. *Am. J. Med. Genet.* 5: 373–381.
- Ségurel, L., M. Wyman, and P. M., 2014 Determinants of mutation rate variation in the human germline. *Annu. Rev. Genomics Hum. Genet.* 15: 47–70.
- Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* 11: 49–78.
- Simons, Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* 46: 220–224.
- Steinberg, A., H. Bleibtreu, T. Kurczynski, A. Martin, and E. Kurczynski, 1967 Genetic studies in an inbred human isolate, pp. 267–290 in *Proceedings of the Third International Congress of Human Genetics*, edited by J. Crow and J. Neel. Johns Hopkins University Press, Baltimore.
- Thomas, C. A., 1971 The genetic organization of chromosomes. *Annu. Rev. Genet.* 5: 237–256.
- White, J. K., A. K. Gerdin, N. A. Karp, E. Ryder, M. Buljan *et al.*, 2013 Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154: 452–464.
- Wiltshire, K. M., R. A. Hegele, A. M. Innes, and A. K. Brownell, 2013 Homozygous lamin A/C familial lipodystrophy R482Q mutation in autosomal recessive Emery Dreifuss muscular dystrophy. *Neuromuscul. Disord.* 23: 265–268.
- Yampolsky, L. Y., F. A. Kondrashov, and A. S. Kondrashov, 2005 Distribution of the strength of selection against amino acid replacements in human proteins. *Hum. Mol. Genet.* 14: 3191–3201.
- Zielenski, J., T. M. Fujiwara, D. Markiewicz, A. J. Paradis, A. I. Anacleto *et al.*, 1993 Identification of the M1101K mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene and complete detection of cystic fibrosis mutations in the Hutterite population. *Am. J. Hum. Genet.* 52: 609–615.

Communicating editor: S. Ramachandran