Received: 30 September 2023 | Accepted: 10 October 2023

DOI: 10.1002/ajmg.a.63452

ORIGINAL ARTICLE

AMERICAN JOURNAL OF medical genetics PART A WILEY

Understanding recessive disease risk in multi-ethnic populations with different degrees of consanguinity

Luis A. La Rocca[1] | Julia Frank[2] | Heidi Beate Bentzen[3] | Jean Tori Pantel[4] | Konrad Gerischer[2] | Anton Bovier[2] | Peter M. Krawitz[1]

[1]Institute for Applied Mathematics, University of Bonn, Bonn, Germany

[2]Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany

[3]Centre for Medical Ethics, Faculty of Medicine, Univeristy of Oslo, Oslo, Norway

[4]Department of Digitalization and General Practice, University Hospital RWTH Aachen, Aachen, Germany

Correspondence
Peter M. Krawitz, Institute for Genomic Statistics and Bioinformatics, University of Bonn, Venusberg-Campus 1, 53127 Bonn, Germany.
Email: pkrawitz@uni-bonn.de

Present address
Julia Frank, Institute of Medical Statistics and Computational Biology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologen, Germany.

Funding information
Deutsche Forschungsgemeinschaft

[Correction added after first online publication on 25 November 2023. Author order has been updated.]

Abstract

Population medical genetics aims at translating clinically relevant findings from recent studies of large cohorts into healthcare for individuals. Genetic counseling concerning reproductive risks and options is still mainly based on family history, and consanguinity is viewed to increase the risk for recessive diseases regardless of the demographics. However, in an increasingly multi-ethnic society with diverse approaches to partner selection, healthcare professionals should also sharpen their intuition for the influence of different mating schemes in non-equilibrium dynamics. We, therefore, revisited the so-called out-of-Africa model and studied in forward simulations with discrete and not overlapping generations the effect of inbreeding on the average number of recessive lethals in the genome. We were able to reproduce in both frameworks the drop in the incidence of recessive disorders, which is a transient phenomenon during and after the growth phase of a population, and therefore showed their equivalence. With the simulation frameworks, we also provide the means to study and visualize the effect of different kin sizes and mating schemes on these parameters for educational purposes.

KEYWORDS
recessive risk counseling, genome-wide carrier screen, non-equilibrium dynamics, Wright–Fisher model

1 | INTRODUCTION

Medical population genetics is dedicated to elucidating the role of genomic variation in susceptibility to diseases and requires expertise in medical genetics, population genetics, epidemiological genetics, and community genetics. This knowledge is usually distributed over many teams and labs and rarely integrated within a single institute, let alone a single person (Giugliani et al., 2019). For the following work, therefore, we imagine a reader who is likely to excel in one of these areas but is only familiar with the foundations of others. We hope that the simulation frameworks we present will be so easy to use that many will end up using them to perform further analysis. In the following, we will motivate the choice of our parameter settings that are based on findings that became available due to recent genome-wide sequencing studies. Sequencing of large cohorts confirmed estimates the number of recessive, lethal equivalents per genome which were previously based on epidemiological data of disease prevalences and stillbirths: On average, healthy individuals carry 0.5–2 heterozygous variants that would prevent reproduction if they occurred in a homozygous state (Bittles & Black, 2010; Chakraborty & Chakravarti, 1977;

Gao et al., 2015; Narasimhan et al., 2016). With respect to population genetics, it is irrelevant whether such variants cause a severe, lethal condition in the affected individual before reproductive age or simply result in complete sterility and are therefore also referred to as lethal equivalents. In simulations that aim at reproducing empirical findings, individuals who are homozygous for a lethal equivalent have a fitness of $s = -1$ and are removed from the gene pool. In contrast, heterozygous carriers of lethal equivalents have the same fitness as wildtypes, $s = 0$, and with respect to simulations, modeling the mating pattern is crucial for the dynamics in population genetics. However, the question of how the ancestral background and the degree of consanguinity affect the recessive lethal load per person is still vividly discussed because empirical data and predictions by theoretical population genetics are partially contradictory (Ballinger & Noor, 2018): in the case of mutation-selection balance, the prevalence of recessive disorders should be the same regardless of ethnicity and mating scheme. However, in the Deciphering Developmental Disorders (DDD) cohort, the proportion of cases due to recessive coding variants was 3.6% in patients of European ancestry, compared to 31% in patients with Pakistani ancestry (Martin et al., 2018). Even within the same population, for example, in Iran, the probability for a recessive cause of intellectual disability is four times higher for offspring from first-cousin unions than for offspring of non-consanguineous partnerships (Hu et al., 2018; Kahrizi et al., 2019; Musante & Ropers, 2014). To explain this discrepancy between the load of recessive lethal variants and the recessive disease burden, some authors recently argued that the unexpectedly high frequency of lethal equivalents might also be explained by an ascertainment bias, that is, some of the pathogenic mutations reached high frequency by chance and are therefore overreported (Amorim et al., 2017). However, since the assumption of mutation-selection balance is not justified, other authors studied the effect of different demographic dynamics including explosive population growth on mutation burden (Henn et al., 2015). Expanding populations incur a mutation burden, also referred to as expansion load, which is a transient phenomenon but can persist for many generations depending on the mating scheme and the coefficients of selection and dominance (Balick et al., 2015; Gravel et al., 2011; Peischl & Excoffier, 2015).

In this work, we explore the influence of different mating schemes in nonequilibrium dynamics by means of two different simulation frameworks with distinct and overlapping generations. Each model had the advantage of handling certain aspects of population genetics particularly well. The first is an adaption of the classical Wright–Fisher model with discrete non-overlapping generations run in the forward genetic simulation framework SLiM (Fisher, 1919; Haller & Messer, 2019). In the second model, generations can overlap because diploid individuals die and give birth at independent exponential times on a continuous timescale (Amorim et al., 2017). For random mating populations with two sexes, the equivalence of the effective population size was already delineated for overlapping generations (Engen et al., 2007). In the following, we show that simulations of the discrete, as well as the overlapping model yield comparable results for an out-of-Africa scenario, suggesting that the existing modeling approaches can be used to fit empirical data that result from nonequilibrium dynamics (Brandvain & Wright, 2016).

2 | METHODS

Editorial policies and ethical considerations. This study was approved as exempt from formal IRB review by Bonn University since no human subject material was analyzed.

Throughout this framework, the mutation burden is defined as the average number of lethal equivalents per individual. The lethal alleles in the genome are deleterious alleles that are disease-causing if both copies of a gene in an individual harbor at least one such variant. The totality of these pathogenic variants could also be regarded as the theoretical superset of an extended carrier screen (Antonarakis, 2019). By this means, we are able to focus on the incidence rate of severe recessive disorders with early onset that prevent reproduction almost with certainty. Likewise, we can study how the selection of a partner, which we refer to as a mating scheme, influences the disease prevalence and mutation burden and we are able to monitor these parameters in the population over time. This is done by counting the number of lethal equivalents that enter the gene pool due to a constant de novo mutation rate, or leave the gene pool due to selection. If the disease prevalence does not change any more, the population is in a steady state, that is a flux balance for lethal equivalents.

In population genetics, the lifespan of individuals that do not reproduce does not matter. In our simulations we therefore used the same age distribution for every individual, regardless of the number of lethal equivalents or the affection status. With the same life span in affected and unaffected individuals, disease prevalence and incidence are also equivalent and their rate is proportional to the amount of lethal equivalents removed from the gene pool per generation or time unit. In fact, the expected number of lethal equivalents that is lost by an affected individual that is not propagating is two. This is equivalent to the difference in the average mutation burden between affected and unaffected individuals and can also be derived from the simulations. An expansion of the population will affect prevalence and mutation burden as we will discuss in more detail in the following.

Consider a finite population of individuals where each individual is characterized by a diploid set of N gene segments of different sizes. Pathogenic variants appear at every gene independently with a rate that is proportional to its size. As long as an individual carries a pathogenic variant at only one gene, its fitness is unaffected. But as soon as both copies of a gene carry a pathogenic variant, the individual's reproductive fitness is reduced to zero. In this case, the individual will be excluded from the mating process and is not able to reproduce any more. Other than that, all individuals are equally fit, no matter how many recessive disorders they carry. Simulations always start with a small, healthy population. After a period of time in which a mutation selection balance is established, a logistic growth phase starts, that settles after a new population equilibrium is reached. We investigate changes of the dynamics of the mutation burden and the prevalence when the population applies different mating schemes. On one hand,

random mating occurs, where individuals select their partner from all potential partners with non-zero fitness uniformly. On the other hand, a consanguineous mating scheme is employed, in which individuals exhibit a preference for mating with close relatives.

2.1 | Discrete model

In the default setting, the simulation package from Haller and Messer (2019) samples a diploid population evolution according to the standard Wright–Fisher model. Sexes were added such that each sex is equally represented in the population at any time. In generation $n \geq 1$ there is a finite number of individuals $M_n \geq 0$ with a total of $2M_n$ genomes alive. In the initial phase the population size is held constant with $M_n = M_0$ for all generations $n \leq n_{\text{grow}}$ in order to establish a mutation selection balance ("burn-in"). Afterwards, the growth phase begins and the population size of each generation grows logistically with growth rate $r > 0$ until it approaches the carrying capacity K. Therefore, the population size of each generation is determined by the following formula

$$M_n = \frac{|K|}{1 + C_0 e^{-rK\left(n - n_{\text{grow}}\right)}} \quad \text{for all } n \geq n_{\text{grow}},$$

where $C_0 = \frac{K - M_0}{M_0}$.

The two mating schemes—random and consanguineous mating—are introduced as following. To generate generation $(n + 1)$, first select M_{n+1} females from generation n independently at random with replacement among all females with non-zero fitness. For the random mating scheme, each female then selects a male uniformly at random from the pool of potential partners who possess positive fitness. To implement the consanguineous mating scheme, utilize the pedigree information provided by SLiM for the last two generations, tracing backwards in time. For each individual, their parents and grandparents are known. In the consanguineous population, a female now selects a mate using a weighted uniform distribution from the set of all potential partners. This choice is influenced by weights α, and $\beta \in [0, 1]$ with $\alpha + \beta \leq 1$. The individual then chooses a male partner with non-zero fitness with

two common grandparents with probability α.

one common grandparent with probability β.

no common grandparents with probability $1 - (\alpha + \beta)$.

Notice that having two grandparents is akin to a cousin relationship, while sharing one grandparent relates to a half-cousins relationship, as depicted in Figure 4a.

To start the simulation select N gene segments from the entire human genome. Each with an independently and uniformly distributed number of base pairs $w_1, \ldots, w_N \sim \cup_{[a,b]}$, where $a, b > 0$, representing the minimum and maximum segment size, respectively. Furthermore, the entire genome is divided into n_c chromosomes. During birth, changes in the offspring's genetic information occur not only through mutation but also via recombination. For each chromosome, initiate an independent *Poisson* Process with rate $r_{\text{rec}} > 0$, which identifies the recombination breakpoints. Here r_{rec} represents the overall recombination rate. The discrete model was implemented in SLiM version 3.2.1.

2.2 | Adaptive dynamics

We employ a diploid version of the adaptive dynamic models introduced by Fournier and Collet (Collet et al., 2013; Fournier & Méléard, 2004). A distinct characteristic of these models lies, firstly, in their foundation on a Poisson process. This entails that individuals produce offspring and undergo mortality at independent rates. Secondly, a noteworthy feature is the ongoing feedback between demographics and ecology due to the competition among individuals. This competitive pressure for finite resources among individuals enables the modeling of a naturally fluctuating population with limited capacity. In the following, we outline the key features of the model. For a comprehensive mathematical description, please refer to Appendix S1. We initiate the simulations within a small, entirely healthy population. This population not only settles into a mutation-selection equilibrium but also experiences fluctuations around a natural population size. This size is contingent upon birth and death rates, as well as the interplay of competition among individuals and the mutation rate. Following the initial burn-in phase, we decrease competition, thereby providing the population with more resources. This alteration triggers logistic population growth until the growth rate tapers off upon reaching the new population equilibrium. To simulate consanguineous mating, we equip each individual additionally to the genetic information with two family flags, aimed at indicating the origin of the individual. During each birth, the newly born individual inherits one randomly chosen family flag from each parent. If both parents possess the same family information, the offspring inherits an identical copy of this information. This modeling approach presents several challenges. Firstly, we must ensure that family groups do not become too large and should periodically disintegrate once the maximum family size of κ is reached. Secondly, this identification mechanism only partially mirrors actual families. For instance, in this model, it is possible that grandparents and their grandchildren do not belong to the same family. In the random mating scheme, individuals select partners randomly from the pool of fit individuals. On the other hand, in the consanguineous mating scheme, partner selection depends on family affiliation. We model the reproductive compatibility between two individuals such that, in an equilibrium population, the probability of selecting a partner with the same family flags is α as long as the family size fluctuates around $\kappa/2$. Conversely, the probability of selecting a partner who shares only one of the family flag with oneself is β. Finally, a partner outside the family is chosen with a probability of $1 - \alpha - \beta$. This holds assuming the population is in equilibrium and the relevant family has a size of $\kappa/2$. During each birth, a Poisson-distributed number of pathogenic variants is randomly distributed across the $2N$ gene segments. The pathogenic variants are allocated to the N genes using a weighted uniform distribution, where the weights correspond to the respective sizes of the genes. Each mutation contributes to the degeneration of the gene segment. There are

no back mutations, beneficial mutations, or neutral mutations in this scenario. Instead of recombination, we employ a form of genetic information reshuffling. During each gamete formation, the genetic information is divided into n_c chromosomes, and from these, one copy is randomly selected. We have implemented the simulations in Python version 3.8 using a Gillespie algorithm.

2.3 | Comparing both models

Both models, the discrete generation model implemented with SLiM and the adaptive dynamics model using the Gillespie algorithm, excel in different aspects of capturing nature. A prominent advantage of SLiM and the discrete model lies in the precise pedigree information generated for every individual. However, the adaptive model can only roughly cluster individuals into family groups and cannot differentiate among members within a single family, as depicted in Figure 4b. Nonetheless, a significant drawback of the discrete model is its non-overlapping generations. This limitation precludes the possibility of matings between individuals on different pedigree levels, such as uncle–niece marriages. This constraint is overcome by the continuous-time model. As individuals independently give birth and die, different generations coexist due to varying ages. The discrete model, similar to the Wright–Fisher model, operates with constant or deterministically increasing population sizes. Conversely, the continuous model accommodates a fluctuating and naturally growing population, as depicted in Figure 5. It is worth noting that for large populations, the random fluctuations in population size are of order K^1, and the stochastic process converges in law to the solution of a deterministic logistic equation (Fournier & Méléard, 2004). Recombination is also approached differently in the two models. SLiM operates with genuine interchromosomal recombination, while the adaptive model simply reshuffles parental chromosomes during gamete production. This distinction arises from SLiM's ability to record the precise base positions of mutations on the human genome. In contrast, the adaptive dynamics model possesses information only about the number of pathogenic variants per gene segment and lacks knowledge of their exact locations within each segment. Given the assumption that all genes are compound heterozygotes, the varying implementations of recombination do not impact the fitness of individuals. However, this reduction in information brings a significant advantage in terms of algorithm runtime.

Apart from all the differences outlined, a substantial effort has been made to ensure parameter equality between both simulations. This includes factors such as the number of gene segments N, the initial and equilibrium population size M_0 and K, and numerous other parameters. Additionally, in the continuous model, family sizes are calibrated to attain an approximate balance between the number of potential partners in the consanguineous setting of both models. Similarly, the birth rate in the continuous-time model is established at $b = 1$, ensuring that within a time interval of $t \in [n, n + 1]$, corresponding to one discrete generation, there are M_{t+1} birth events, where M_t denotes the population size at that particular time. The only distinction lies in the discrete generation model having *exactly* M_{n+1} births in generation n, whereas the continuous-time model experiences *on average* that number of births.

3 | RESULTS

We initiate our simulations with a population of 500 individuals, allowing for approximately 500 generations to reach a steady state, that is no significant change in the mutation burden. A comparable size has also been suggested for the population that left the African continent 10,000 to 200,000 years ago (Gutenkunst et al., 2009; Tenesa et al., 2007). Following this out-of-Africa event, the population expands to a size of 10,000 individuals in approximately 130 generations. This corresponds to an estimated duration of around 2500 years and an average growth of 1–2% per generation. The population expansion adheres to a logistic growth curve, which takes on the appearance of a step function (as depicted by the gray curve in Figure 1), due to the extensive duration of our complete simulations spanning 2000 generations. All individuals have diploid genomes with 1000 recessive genes that we deem crucial for reproductive success. Their coding sequence ranges between 500 and 10,000 base pairs (bp) per gene, novel alleles are introduced with a de novo mutation rate of 1.2×10^{-8} per bp, and one out of nine mutations is expected to be a lethal equivalent (Kong et al., 2002, 2012). The choice of these parameters are motivated by the distribution of coding lengths and the deleteriousness scores for known autosomal recessive genes (Kircher et al., 2014; Kochinke et al., 2016). Pairs for procreation are formed either randomly or based on their relatedness that is traced over the two most recent generations. In a highly consanguineous mating scheme, the number of potential partners is hardly affected by the population size, as most marriages happen within families. In our simulations, this mating scheme is realized as follows: 50% of all partnerships share two grandparents, 30% share one grandparent, and only 20% share no grandparent. In this scenario, the mutation burden and prevalence do not change during population growth (Figure 1b). However, linkage disequilibrium suggests that out-of-Africa populations have only reached effective populations sizes of around 3k, thus this might be an overestimate (Tenesa et al., 2007). In contrast, in a randomly mating population, there is a sharp transient drop of incidence rates during expansion at the expense of an increasing mutation burden (Figure 1a). However, after the final size of the population is reached, it takes almost another 550 generations until the mutation burden reaches its new plateau of approximately three pathogenic variants in 1000 recessive disease genes. In contrast to the mutation burden, the prevalence is independent of effective population size and a function of mutation rate only. For constant μ, the prevalence returns to the initial value before the expansion. Since affected individuals in our simulation have the same life expectancy and only do not propagate, prevalence and incidence are the same and there are roughly 70 affected individuals per generation in a population of 10,000 or 0.7%. The mutation burden in the steady state increases in both mating schemes with the number of autosomal recessive genes,

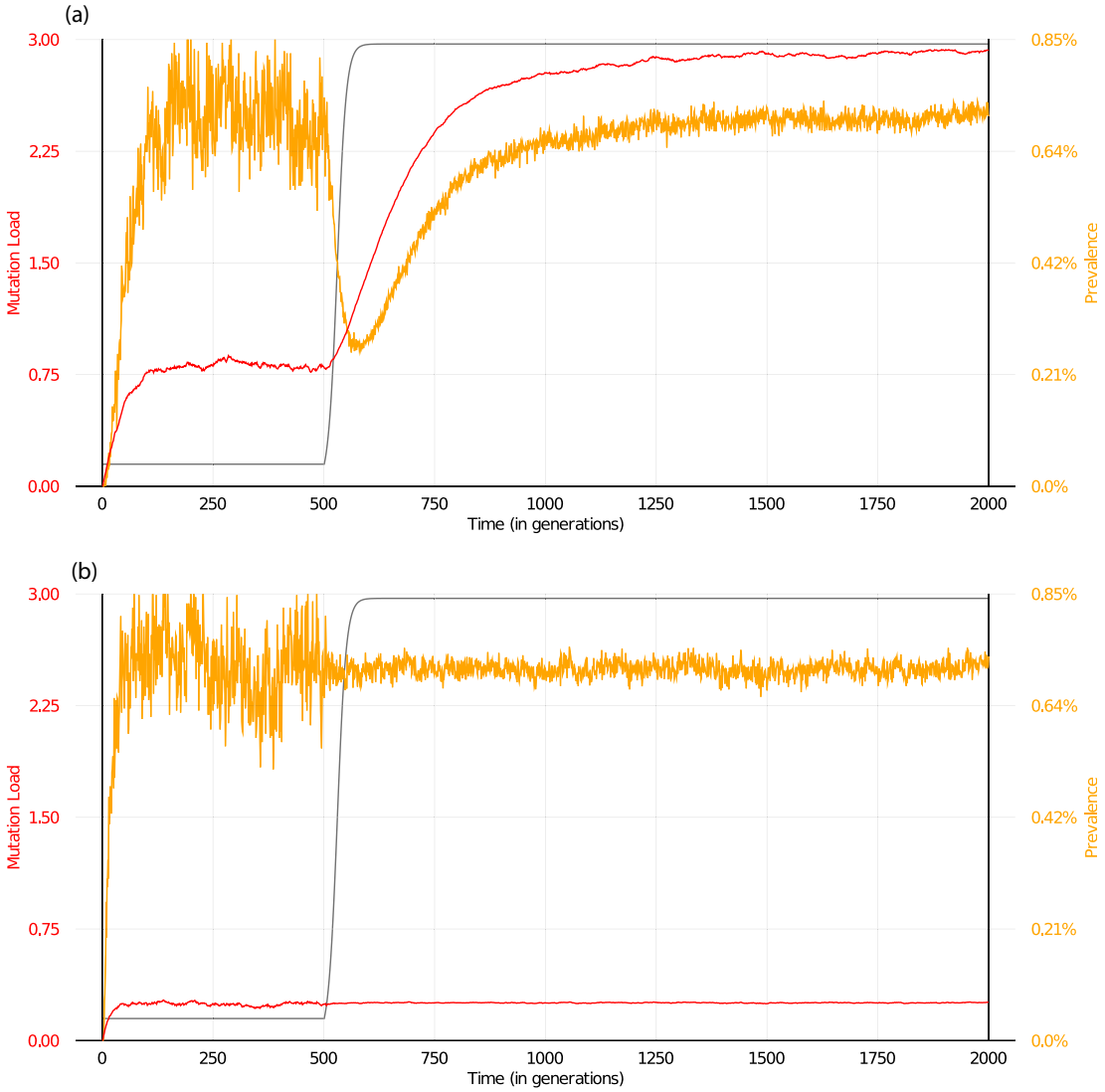

FIGURE 1 Dynamics of mutation load and prevalence for severe recessive disorders: (a) population expansion from 500 to 10,000 individuals (gray), starting in generation 500 does not affect prevalence (orange) nor mutation load (red) if partners are preferentially chosen within relatives (consanguineous mating scheme) (b). In contrast, in a random mating population, there is a transient drop of prevalence at the expense of an increasing mutation load (a). It takes more than 550 generations after the end of the growth phase, until the steady state is reached and the prevalence for both mating schemes are comparable again. The plots show the average of 50 exact trajectories of the stochastic process simulated with the Wright–Fisher model.

but with population size only for random mating (Figure 3a,b). This is best explained by a limit of the effective number of available partners that the consanguineous mating scheme imposes, regardless of the final population size. In line with that argument, there is a transition from the dynamics of consanguineous to random when we incrementally increase family size, which would correspond to more potential mating partners (Figure 3). Although the phase of population growth lasts only 130 generations in our simulations, the time span to reach the new equilibrium for the mutation burden lasts much longer. In both simulation frameworks, we were able to achieve numbers of lethal equivalents that are in accordance with observations from the literature that are based on epidemiological data as well as population genetic data. In a recent study, Narasimhan et al. analyzed exomes of 3222 British adults of Pakistani heritage with a high parental relatedness and found a significantly lower number of homozygous knockout genotypes than expected from the summary statistics of a more outbred population. By this means, they were able to compute an average number of 1.6 recessive-lethal equivalents per individual (Narasimhan et al., 2016). In mutation-selection balance, the number of recessive-lethal equivalents is only a function of genome architecture and the effective population, which the mating scheme influences. In the non-equilibrium dynamics, however, the choice of the partner has the greatest influence on the increase of recessive lethal equivalents. Since human societies almost mirror the unmanageable variety of mating systems in the mammalian kingdom it is noteworthy that with the discrete and adaptive simulations, different aspects of

mating can be modeled (Clutton-Brock, 1989). In the adaptive framework, for example, we allowed partnerships between different generations and for each offspring the parents were selected anew (lottery polygyny) (Caballero, 1994). Despite the differences in the implementation details, both simulations yielded comparable dynamics when the extended family size κ and the autozygosity were adjusted. Over certain historic periods, the extended family size κ, which was the parameter used in the adaptive model, might be easier to delineate. Whereas kinship coefficients could be estimated with exact pedigrees and genomic data. We therefore extended the possibilities of how empirical data can be explained by population genetic simulations.

4 | DISCUSSION

The empirical observation that consanguinity is associated with an increased risk of autosomal recessive disorders, has been made in many countries but are only based on records of relatively few generations. Martin et al. showed that the contribution of autosomal recessive developmental disorders is 31% in the current British population if the autozygosity is above 0.02 (Martin et al., 2018). Likewise, in the Iranian population it is estimated that offspring from first-cousin unions have a probability for intellectual disabilities that is four times higher than in non-consanguineous partnerships (Hu et al., 2018; Kahrizi et al., 2019; Musante & Ropers, 2014). Although population genetics predicts these findings as a transient phenomenon in nonequilibrium dynamics, this literature is often not cited in the empirical works (Balick et al., 2015; Glémin, 2003; Gravel et al., 2011; Henn et al., 2015; Kirkpatrick & Jarne, 2000; Lohmueller et al., 2008; Peischl & Excoffier, 2015; Simons & Sella, 2016). In our work we studied how rapid changes in population size affect the expected number of lethal equivalents when generations overlap, and achieved similar results as in the Wright–Fisher model. By that means we addressed an outstanding question in nonequilibrium population genetics. We hypothesize that epidemiological data accumulated over a few centuries, which is a short time period with respect to recessive selection and a lack of knowledge in population genetics, might frame a biased risk perception that might even influence aspects of social norms. According to our simulations and previous work, the advantage of outbreeding is a transient phenomenon for a population that is initially in mutation-selection balance and that starts to grow. The lower prevalence compared to an inbred population lasts for many generations even after the expansion phase has ended, until mutation-selection balance is reached again with a higher count of lethal equivalents. We found it intriguing that, for example, first-cousin marriage in Europe was banned after several generations of population growth during the Roman empire and considerable migration and admixture (Henn et al., 2015). While this continent clearly benefitted at that time point from a change of social conventions with respect to the recessive disease burden, the consequences of different mating schemes, for example, on the proportion of congenital malformation are less prominent in populations that were more constant in size over a long period of time (Chakraborty & Chakravarti, 1977). One of the most extreme examples of descendants of a small group might be the Hutterites, who increased in population size by more than a factor of 400 in less

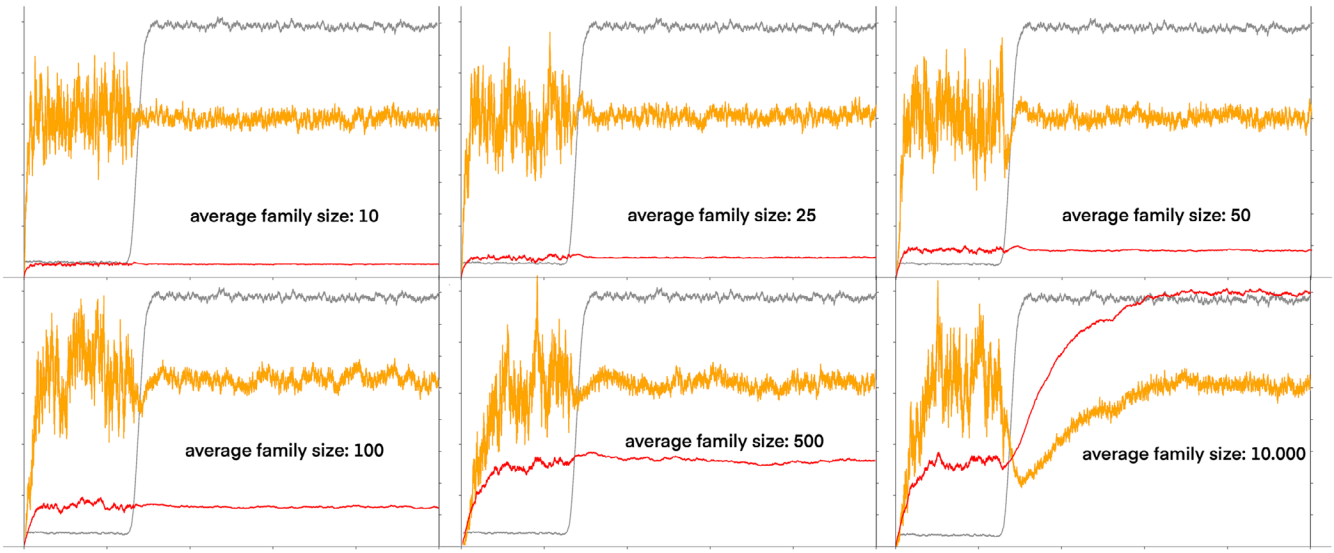

FIGURE 2 Influence of family size on mutation load and prevalence: The mating scheme is characterized by the family size and a probability function that describes how many of the partners are chosen within the family. In a preferentially consanguineous mating population the dynamics change when the maximum family size increases (upper left panel to lower right from 10, 25, 50, 100, 500 up to 10,000). The mutation load starts to increase considerably if mating is happening in tribes of 500 individuals. However, at this stage there is still only a minor effect of further population growth. In the lower right the maximum of the allowed family size is equivalent to the population size and thus, dynamics do not differ from a random mating scheme any more. The plots show the average of 10 exact trajectories of the stochastic process simulated with the individual-based model of adaptive dynamics model.

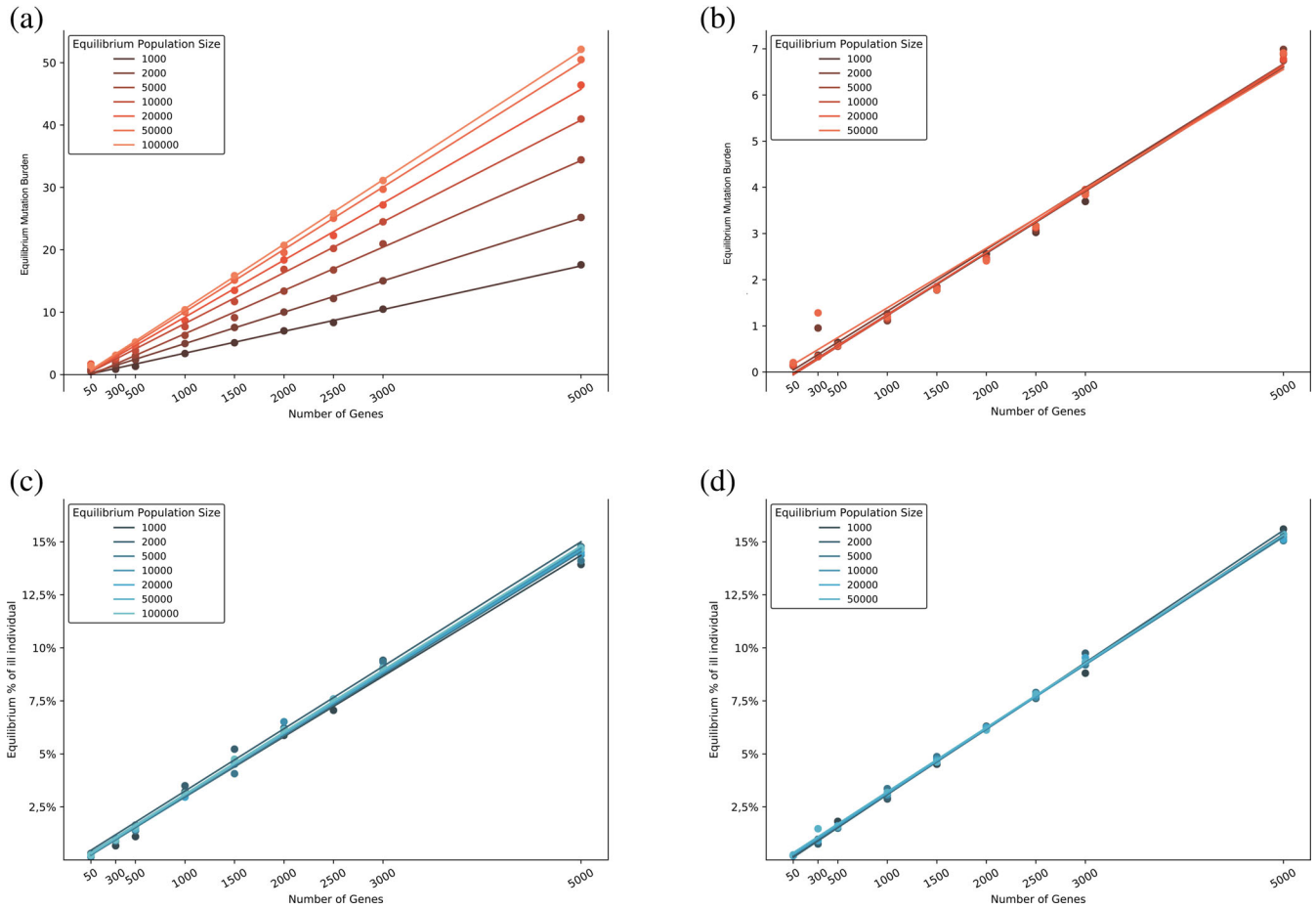

FIGURE 3 Influence of genomic architecture and population size: The capacity of the genome for deleterious mutations is larger in the random mating population. With an increasing number of genes and growing population size, deleterious mutations accumulate (a). In contrast, in the consanguineous mating scheme, family size limits the effective population size, and therefore mutation load is independent of the total number of individuals (b). Prevalence increases linearly in both mating schemes when the number of genes increases and is independent from population size, as regression analysis indicates (c,d).

than 200 years from a founding population of less than 100 people (Boycott et al., 2008). This is comparable to a kin of 100 which is still too small to benefit from a drop in prevalence during growth as shown in Figure 2. The few initial lethal equivalents of the founders were amplified to high prevalence and are now also listed as recessive alleles of high frequency in the database of genetic disorders in Amish, Mennonite and Hutterite (Payne et al., 2011). However, a transient reduction of recessive disease burden can be achieved by marriage that is colony exogamous, which is also most likely for that reason a social accepted mating scheme. The occurrence and coexistence of different marriage patterns over many centuries can certainly not be understood by population genetics alone since social, cultural and economic factors interact with demographics in a complex manner (Henn et al., 2015). It is therefore concerning when questionable genetic reasoning is used in the legislature. For instance, the European Court of Human Rights case of *Stübing v. Germany* concerned consanguineous siblings who had four children following consensual intercourse, whereupon both siblings were charged with incest ("Stübing v. Germany," 2012). One of the siblings lodged a complaint, arguing that the legislature violated his right to sexual self-determination, his private and family life. The Court found that 24 out of 44 European States reviewed, criminalized consensual sexual acts between adult siblings, and all prohibited siblings from getting married. The German government argued that the law against incest partly aimed to protect against the significantly increased risk of genetic damage among children from an incestuous relationship ("Beschluss des Zweiten Senats," 2008). However, motivating a law to avoid a higher probability of disease can be viewed as eugenic: As the German Ethics Council opined after the judgment, no convincing argument can be derived from there being a risk of genetic damage (Deutscher Ethikrat, 2014). The Council also pointed out that prohibiting procreation in non-consanguineous couples who carry a genetic burden, would not be allowed to be proposed or considered in any manner (Deutscher Ethikrat, 2014). Any prohibition of consanguineous relationships should therefore build on non-genetic reasoning. The view of the German Ethics Council concurs with a statement by the German Society of Human Genetics criticizing eugenic reasoning in a judgment by the German Federal Constitutional Court in 2008 on

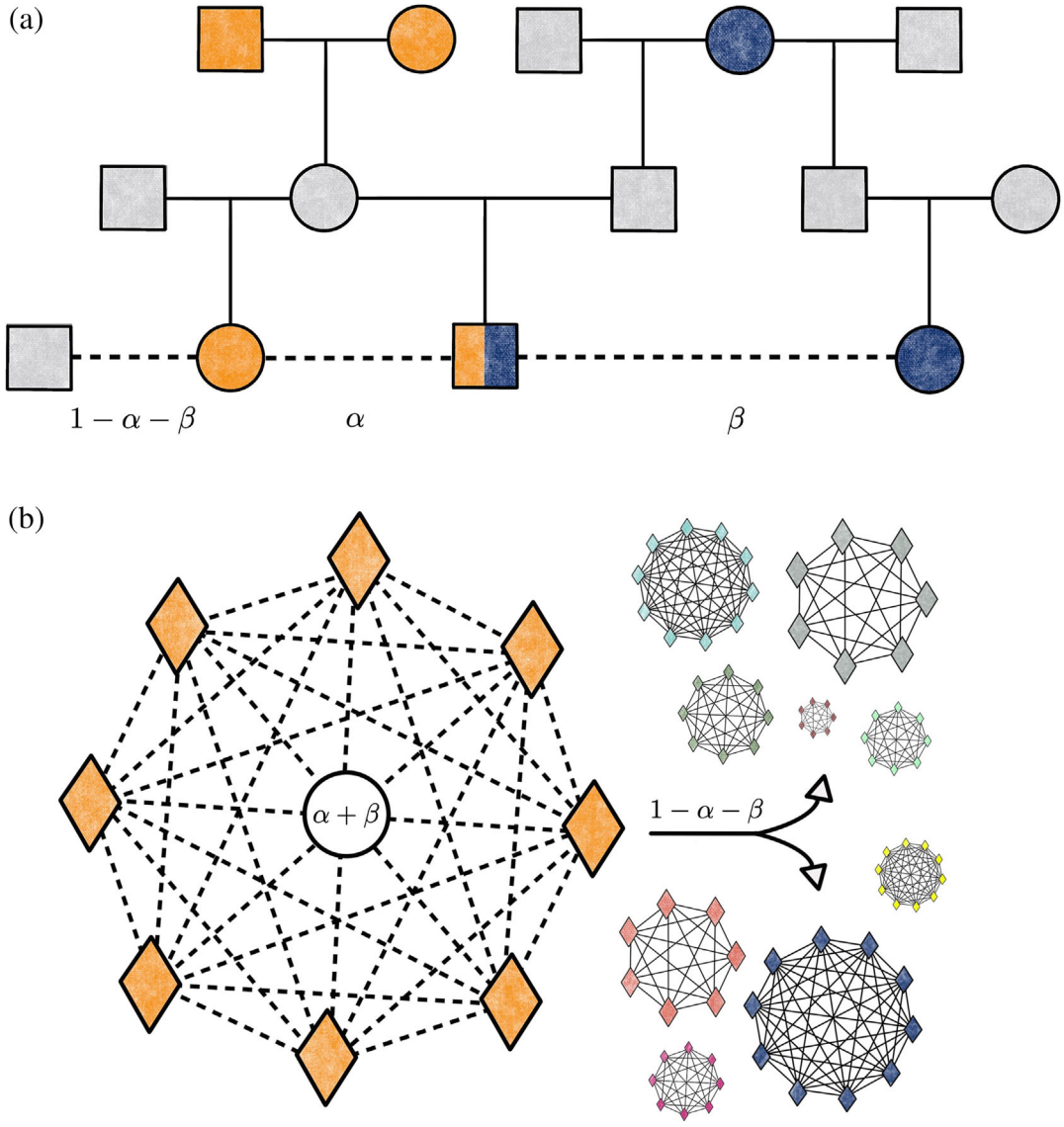

FIGURE 4 Comparison of implementation of consanguineous mating scheme: The upper image (a) depicts a typical pedigree resulting from the implementation of consanguineous mating in SLiM. Precise inheritance up to two generations in the past are known. it is highly unlikely for two parents to have more than one child together since females independently choose partners for each mating. The pivotal factor in partner selection is the number of shared ancestors in the previous generation lineage. If two parents have two common ancestors, a mating occurs with probability α; if they share one common ancestor from two generations ago, mating occurs with probability β; and if they lack any common ancestors, mating transpires with probability $1-\alpha-\beta$. In the lower image (b), a schematic visualization of consanguineous mating in the adaptive model is presented. Within families, no specific structures is retained. Mating within the family occurs with a probability of $\alpha+\beta$, while mating outside the family transpires with a probability of $1-\alpha-\beta$. In addition to the probabilities α, β the average family size $\kappa/2$ plays a decisive role here.

criminal liability of incest between siblings. The Society stated that "The argument that reproduction needs to be thwarted in couples whose children possess an elevated risk for recessively inherited illnesses is an attack on the reproductive freedom of all." ["Das Argument, es müsse in Partnerschaften, deren Kindern ein erhöhtes Risiko für rezessive erbliche Krankheiten haben, einer Fortpflanzung entgegengewirkt werden, ist ein Angriff auf die reproduktive Freiheit aller".] ("Stellungnahme der Deutschen Gesellschaft für Humangenetik (GfH)," 2008). The Society added that apart from being factually incorrect, eugenic reasoning also encourages discrimination and should therefore be avoided by the courts ("Stellungnahme der Deutschen Gesellschaft für Humangenetik (GfH)," 2008).

Furthermore, as our work shows, the argument that there exists an increased risk of genetic damage requires the definition of a reference population for comparison. However, there is neither agreement about a suitable reference nor an accurate measurement for mutation burden (Henn et al., 2015). When genetic counseling is sought, the predicted recessive disease burden that is communicated in the consultation might influence decisions, for example, about the choice of partners or family planning. Since this risk does not only depend on

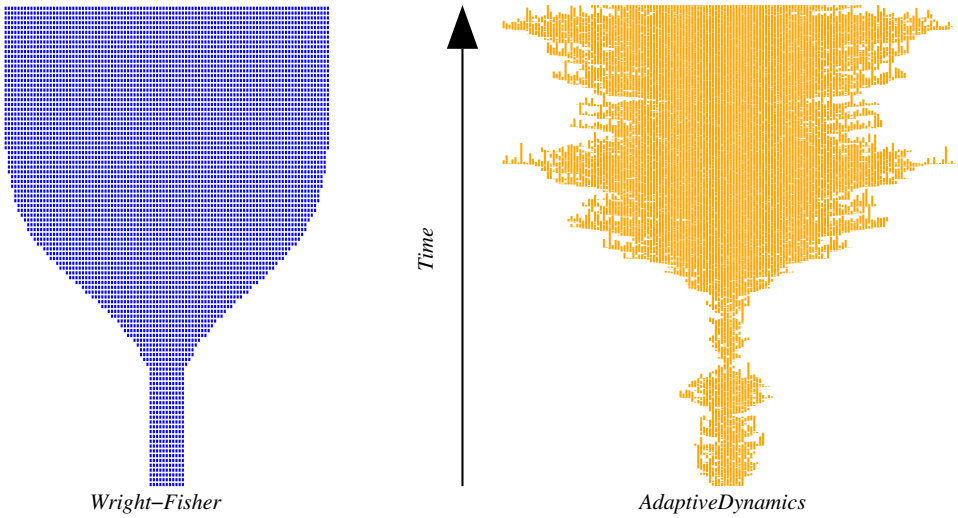

FIGURE 5 Comparison of population size and life spans of individuals: Each tick represents the lifespan of an individual. Time progresses from bottom to top. On the left side in blue, one can observe that all individual in the Wright–Fisher model share the same lifespan, generations do not overlap, and the population is of constant size initially, then grows deterministically until the new constant size is reached. In contrast, on the right side in orange, the adaptive model exhibits varying lifespans among individuals, leading to population fluctuations. Moreover, birth times of individuals are independent of each other, resulting in smoothly transitioning generations. At the point where the deterministic growth starts in the Wright–Fisher model the adaptive dynamics population was given more capacity which also leads to a logistic grow.

mating schemes but also on mutation burden it is important to measure this parameter as accurately as possible. In our simulations, an individual of the outbreed population had on average four times more lethal equivalents than an individual of the inbred population when the mutation-selection balance was reached again many generations after the growth phase ended.

Interestingly, these values and the range are comparable to what has also been described in the literature for real populations. With respect to the British subpopulations of Pakistani (PABI) and European (EABI) ancestry in Martin et al., this could mean that PABI with a considerably higher autozygosity and many first-cousin marriages are closer to mutation-selection balance than EABI. This would imply that the disease prevalence for recessive disorders will remain constant for PABI while it will approach that level for EABI in the following generations, given that the different mating schemes continue. In contrast, the higher mutation burden in the EABI subgroup due to the higher effective population size might already now contribute to a higher risk for autism spectrum disorders, which are also highly heritable but do not follow monogenic inheritance (Ji et al., 2016). Since assessing recessive lethals based on family history is very challenging, genetic counseling should increasingly focus on carrier testing in cases where individuals seek help to gain information to make their own decisions. Based on current ClinVar statistics, there are more than 150,000 pathogenic alleles known for recessive genes that cause severe disorders. In large German cohort of individuals with rare disorders, a diagnosis could be established in 125 cases due to homozygosity or compound heterozygosity of pathogenic variants. Ninety-four of these causative variants would also have been classified as pathogenic in the healthy parents in a preconceptional exome analysis (Schmidt et al., 2023). Expanded carrier screens can play an important role in genetic counseling in multi-ethnic populations with different degrees of consanguinity, and it should be discussed who should have access to this test to make their own informed decisions (Antonarakis, 2019; Schmidtke & Cornel, 2020).

AUTHOR CONTRIBUTIONS

Anton Bovier and Peter M. Krawitz conceptualized the study. Luis A. La Rocca, Julia Frank and Konrad Gerischer conducted the simulations and analyzed the data. Heidi Beate Bentzen assessed the legal and ethical implications. Jean Tori Pantel visualized the data. All authors wrote and reviewed the manuscript.

ACKNOWLEDGMENTS

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy GZ 2047/1, Projekt-ID 390685813 and GZ 2151, ProjectID 390873048 and through the Priority Programme 1590 "Probabilistic Structures in Evolution". We thank Benjamin C. Haller for his very helpful advices on how to best use the forward genetic simulation framework SLiM (Haller & Messer, 2019).

DATA AVAILABILITY STATEMENT

All scripts to reproduce our simulation results can be found in the following repository: https://github.com/roccminton/Diploid_Model_Two_Loci A video clip of our simulations can be found at: https://youtu.be/5hOgLyRqWPg

ORCID

Luis A. La Rocca https://orcid.org/0000-0001-8523-556X
Heidi Beate Bentzen https://orcid.org/0000-0001-8285-818X

Jean Tori Pantel https://orcid.org/0000-0002-2674-4660
Anton Bovier https://orcid.org/0000-0003-1612-2056
Peter M. Krawitz https://orcid.org/0000-0002-3194-8625

REFERENCES

Amorim, C. E. G., Gao, Z., Baker, Z., Diesel, J. F., Simons, Y. B., Haque, I. S., Pickrell, J., & Prze-worski, M. (2017, September). The population genetics of human disease: The case of recessive, lethal mutations. *PLoS Genetics*, *13*(9), e1006915.

Antonarakis, S. E. (2019, September). Carrier screening for recessive disorders. *Nature Reviews. Genetics*, *20*(9), 549–561.

Balick, D. J., Do, R., Cassa, C. A., Reich, D., & Sunyaev, S. R. (2015, August). Dominance of deleterious alleles controls the response to a population bottleneck. *PLoS Genetics*, *11*(8), e1005436.

Ballinger, M. A., & Noor, M. A. F. (2018, February). Are lethal alleles too abundant in humans? *Trends in Genetics*, *34*(2), 87–89.

Beschluss des Zweiten Senats. (2008, February). Bundesverfassungsgericht. 2 BvR 392/07. Retrieved from https://www.bverfg.de/e/rs20080226_2bvr039207.html

Bittles, A. H., & Black, M. L. (2010, January). Evolution in health and medicine sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(Suppl 1), 1779–1786.

Boycott, K. M., Parboosingh, J. S., Chodirker, B. N., Lowry, R. B., McLeod, D. R., Morris, J., Greenberg, C. R., Chudley, A. E., Bernier, F. P., Midgley, J., Møller, L. B., & Innes, A. M. (2008, April 15). Clinical genetics and the Hutterite population: A review of Mendelian disorders. *American Journal of Medical Genetics A*, *146A*(8), 1088–1098. https://doi.org/10.1002/ajmg.a.32245

Brandvain, Y., & Wright, S. I. (2016, April). The limits of natural selection in a Nonequilibrium world. *Trends in Genetics*, *32*(4), 201–210.

Caballero, A. (1994, December 1). Developments in the prediction of effective population size. *Heredity*, *73*(6), 657–679. https://doi.org/10.1038/hdy.1994.174

Chakraborty, R., & Chakravarti, A. (1977, April). On consanguineous marriages and the genetic load. *Human Genetics*, *36*(1), 47–54.

Clutton-Brock, T. H. (1989, May 22). Mammalian mating systems. *Proceedings of the Royal Society of London - Series B: Biological Sciences*, *236*(1285), 339–372. https://doi.org/10.1098/rspb.1989.0027

Collet, P., Méléard, S., & Metz, J. A. J. (2013). A rigorous model study of the adaptive dynamics of Mendelian diploids. *Journal of Mathematical Biologie*, *67*, 569–607.

Deutscher Ethikrat. (2014). Incest prohibition. Opinion ISBn 978-3-941957-61-9. Retrieved from https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/englisch/opinion-incest-prohibition.pdf

Engen, S., Ringsby, T. H., Saether, B.-E., Lande, R., Jensen, H., Lillegärd, M., & Ellegren, H. (2007, August). Effective size of fluctuating populations with two sexes and overlapping generations. *Evolution*, *61*(8), 1873–1885.

Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *52*(2), 399–433.

Fournier, N., & Méléard, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Annals of Applied Probability*, *14*(4), 1880–1919.

Gao, Z., Waggoner, D., Stephens, M., Ober, C., & Przeworski, M. (2015, April). An estimate of the average number of recessive lethal mutations carried by humans. *Genetics*, *199*(4), 1243–1254.

Giugliani, R., Bender, F., Couto, R., Bochernitsan, A., Brusius-Facchin, A. C., Burin, M., Amorim, T., Acosta, A. X., Purificação, A., Leistner-Segal, S., Saraiva-Pereira, M. L., Jardim, L. B., Matte, U., Riegel, M., Cardoso-dos-Santos, A. C., Rodrigues, G., Oliveira, M. Z., Tagliani-Ribeiro, A., Heck, S., … Kubaski, F. (2019). Population medical genetics: Translating science to the community. *Genetics and Molecular Biology*, *42*(1 Suppl 1), 312–320. https://doi.org/10.1590/1678-4685-GMB-2018-0096

Glémin, S. (2003, December). How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution*, *57*(12), 2678–2687.

Gravel, S., Henn, B. M., Gutenkunst, R. N., Indap, A. R., Marth, G. T., Clark, A. G., Yu, F., Gibbs, R. A., 1000 Genomes Project, & Bustamante, C. D. (2011, July). Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 11983–11988.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009, October). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), 1–11. https://doi.org/10.1371/journal.pgen.1000695

Haller, B. C., & Messer, P. W. (2019, March). SLiM 3: Forward genetic simulations beyond the Wright-Fisher model. *Molecular Biology and Evolution*, *36*(3), 632–637.

Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G., & Gravel, S. (2015, June). Estimating the mutation load in human genomes. *Nature Reviews. Genetics*, *16*(6), 333–343.

Hu, H., Kahrizi, K., Musante, L., Fattahi, Z., Herwig, R., Hosseini, M., Oppitz, C., Abedini, S. S., Suckow, V., Larti, F., Beheshtian, M., Lipkowitz, B., Akhtarkhavari, T., Mehvari, S., Otto, S., Mohseni, M., Arzhangi, S., Jamali, P., Mojahedi, F., … Najmabadi, H. (2018, January). Genetics of intellectual disability in consanguineous families. *Molecular Psychiatry*, *24*, 1027–1039.

Ji, X., Kember, R. L., Brown, C. D., & Bućan, M. (2016, December). Increased burden of deleterious variants in essential genes in autism spectrum disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(52), 15054–15059.

Kahrizi, K., Hu, H., Hosseini, M., Kalscheuer, V. M., Fattahi, Z., Beheshtian, M., Suckow, V., Mohseni, M., Lipkowitz, B., Mehvari, S., Mehrjoo, Z., Akhtarkhavari, T., Ghaderi, Z., Rahimi, M., Arzhangi, S., Jamali, P., Chian, M. F., Nikuei, P., Kermani, F. S., … Ropers, H.-H. (2019, January). Effect of inbreeding on intellectual disability revisited by trio sequencing. *Clinical Genetics*, *95*(1), 151–159.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014, March). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315.

Kirkpatrick, M., & Jarne, P. (2000, February). The effects of a bottleneck on inbreeding depression and the genetic load. *The American Naturalist*, *155*(2), 154–167.

Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M. A. W., Kleefstra, T., Kramer, J. M., Webber, C., Huynen, M. A., & Schenck, A. (2016, January). Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *American Journal of Human Genetics*, *98*(1), 149–164.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S. W., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., … Stefansson, K. (2012, August). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*(7412), 471–475. https://doi.org/10.1038/nature11396

Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., & Stefansson, K. (2002, July). A high-resolution recombination map of the human genome. *Nature Genetics*, *31*(3), 241–247. https://doi.org/10.1038/ng917

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., Sninsky, J. J., White, T. J., Sunyaev, S. R., Nielsen, R., Clark, A. G., & Bustamante, C. D. (2008, February). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, *451*(7181), 994–997.

Martin, H. C., Jones, W. D., McIntyre, R., Sanchez-Andrade, G., Sanderson, M., Stephenson, J. D., Jones, C. P., Handsaker, J., Gallone, G., Bruntraeger, M., McRae, J. F., Prigmore, E., Short, P., Niemi, M., Kaplanis, J., Radford, E. J., Akawi, N., Balasubramanian, M., Dean, J., … Deciphering Developmental Disorders Study. (2018, December). Quantifying the contribution of recessive coding variation to developmental disorders. *Science*, *362*(6419), 1161–1164.

Musante, L., & Ropers, H. H. (2014, January). Genetics of recessive cognitive disorders. *Trends in Genetics*, *30*(1), 32–39.

Narasimhan, V. M., Hunt, K. A., Mason, D., Baker, C. L., Karczewski, K. J., Barnes, M. R., Barnett, A. H., Bates, C., Bellary, S., Bockett, N. A., Giorda, K., Griffiths, C. J., Hemingway, H., Jia, Z., Kelly, M. A., Khawaja, H. A., Lek, M., McCarthy, S., McEachan, R., … van Heel, D. A. (2016, April). Health and population effects of rare gene knockouts in adult humans with related parents. *Science*, *352*(6284), 474–477.

Stellungnahme der Deutschen Gesellschaft für Humangenetik (GfH). (2008). Eugenische Argumentation im Beschluss des Bundesverfassungsgerichts zum Inzestverbot. *Medizinische Genetik*, *20*(2), 239. https://doi.org/10.1007/s11825-008-0099-6

Stübing v. Germany. (2012, April). European court of human rights. App no. 43547/08.

Payne, M., Rupar, C. A., Siu, G. M., & Siu, V. M. (2011, March). Amish, mennonite, and hutterite genetic disorder database. *Paediatrics & Child Health*, *16*(3), e23–e24. https://doi.org/10.1093/pch/16.3.e23

Peischl, S., & Excoffier, L. (2015, May). Expansion load: Recessive mutations and the role of standing genetic variation. *Molecular Ecology*, *24*(9), 2084–2094.

Schmidt, A., Danyel, M., Grundmann, K., Brunet, T., Klinkhammer, H., Hsieh, T.-C., Engels, H., Peters, S., Knaus, A., Moosa, S., Averdunk, L., Boschann, F., Sczakiel, H., Schwartzmann, S., Mensah, M. A., Pantel, J. T., Holtgrewe, M., Bösch, A., Weiß, C., … Wagner, M. (2023, April). *Next-generation phenotyping integrated in a national framework for patients with ultra-rare disorders improves genetic diagnostics and yields new molecular findings*.

Schmidtke, J., & Cornel, M. C. (2020, January). Contentious ethical issues in community genetics: let's talk about them. *Journal of Community Genetics*, *11*(1), 5–6. https://doi.org/10.1007/s12687-019-00444-5

Simons, Y. B., & Sella, G. (2016, December). The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Current Opinion in Genetics & Development*, *41*, 150–158.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., & Visscher, P. M. (2007, April). Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, *17*(4), 520–526. https://doi.org/10.1101/gr.6023607

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: La Rocca, L. A., Frank, J., Bentzen, H. B., Pantel, J. T., Gerischer, K., Bovier, A., & Krawitz, P. M. (2023). Understanding recessive disease risk in multi-ethnic populations with different degrees of consanguinity. *American Journal of Medical Genetics Part A*, 1–11. https://doi.org/10.1002/ajmg.a.63452