



A test for deviations from expected genotype frequencies on the X chromosome for sex-biased admixed populations

Daniel Backenroth¹ · Shai Carmi¹

Received: 9 February 2019 / Revised: 19 April 2019 / Accepted: 29 April 2019 / Published online: 17 May 2019
© The Genetics Society 2019

Abstract

Genome-wide scans for deviations from expected genotype frequencies, as determined by the Hardy–Weinberg equilibrium (HWE), are commonly applied to detect genotyping errors and deviations from random mating. In contrast to the autosomes, genotype frequencies on the X chromosome do not reach HWE within a single generation. Instead, if allele frequencies in males and females initially differ, they oscillate for a few generations toward equilibrium. Allele frequency differences between the sexes are expected in populations that have experienced recent sex-biased admixture, namely, their male and female founders differed in ancestry. Sex-biased admixture does not allow testing for HWE on X, because deviations are *naturally* expected, even under random mating (post admixture) and error-free genotyping. In this paper, we develop a likelihood ratio test and a χ^2 test to detect deviations from expected genotype frequencies on X, beyond natural deviations due to sex-biased admixture. We demonstrate by simulations that our tests are powerful for detecting deviations due to non-random mating, while at the same time they do not reject the null under historical sex-biased admixture and random mating thereafter. We also demonstrate that when applied to 1000 Genomes project populations, our likelihood ratio test rejects fewer SNPs than other tests, but we describe limitations in the interpretation of the results.

Introduction

Testing for deviations from Hardy–Weinberg equilibrium (HWE) is an important quality control step in genome-wide association studies (Anderson et al. 2010; Bycroft et al. 2018; Laurie et al. 2010; Turner et al. 2011). Extensive literature exists on HWE tests for the autosomes, from classical tests to recent work on Bayesian approaches, structured populations, sequenced and imputed genotypes, and software tools (Ayres and Balding 1998; Bourgain et al. 2004; Emigh 1980; Graffelman 2015; Graffelman et al. 2017; Graffelman et al. 2013; Hao and Storey 2017; Hernandez and Weir 1989; Levene 1949; Rohlfs and Weir 2008; Shriner 2011; Wakefield 2010; Wigginton et al. 2005; Yu et al. 2009). However, tests for HWE on the X chromosome have only been recently developed (Graffelman and Weir 2016, 2018; Puig et al. 2017; You et al. 2015;

Zheng et al. 2007). The importance of associations of X-linked variants with complex traits, particularly as a mechanism of sexual dimorphism, has been recently recognized (Chang et al. 2014; Gao et al. 2015; Khramtsova et al. 2019; Kudelka et al. 2016; Kukurba et al. 2016; Li et al. 2015; Scelsi et al. 2018; Traglia et al. 2017; Yap et al. 2018), and these developments underscore the importance of proper quality control on X, including testing for deviations from HWE.

A naive test for HWE on X would consider females only. However, such a test would implicitly assume an equal allele frequency between males and females. Indeed, a number of tests have been recently proposed for *joint* testing of HWE in females as well as equality of allele frequencies between the sexes (Graffelman and Weir 2016; Puig et al. 2017; You et al. 2015). However, these tests ignore the possibility that allele frequencies in males and females would differ *naturally* due to sex-biased admixture.

While autosomal allele frequencies reach HWE within a single generation, it is well known that for X, in case male and female allele frequencies initially differ, equilibrium (in an infinite population) is only asymptotically reached (Jennings 1916; Rosenberg 2016). The classical equations describing the evolution of allele frequencies on X, for an

✉ Shai Carmi
shai.carmi@huji.ac.il

¹ Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

infinite population, are,

$$p_f(t + 1) = \frac{p_m(t) + p_f(t)}{2}, \tag{1}$$

$$p_m(t + 1) = p_f(t), \tag{2}$$

where $p_f(t)$ and $p_m(t)$ are the female and male allele frequencies, respectively, at generation t . Starting with unequal allele frequencies at generation $t = 0$, the male and female frequencies oscillate while gradually stabilizing. Specifically (Rosenberg 2016), if $p_f(0) = 1$ and $p_m(0) = 0$, then

$$p_f(t) = \frac{2^{t+1} + (-1)^t}{3 \cdot 2^t} \tag{3}$$

$$p_m(t) = \frac{2^t - (-1)^t}{3 \cdot 2^{t-1}}. \tag{4}$$

While equilibrium is approached exponentially quickly, if allele frequencies initially differ by a substantial amount, the frequency difference between the sexes can be non-negligible in the first few generations.

Recent sex-biased admixture has been identified for several populations, in particular in the Pacific and the Americas (Bonnen et al. 2010; Bryc et al. 2010; Jagadeesan et al. 2018; Kim et al. 2012; Lie et al. 2007; Lind et al. 2007; Mathias et al. 2016). Moreover, admixture in these populations has often been cross-continental, which may have led to large initial frequency differences between the sexes. Thus, even if a population has been randomly mating since admixture, and even if SNPs are accurately genotyped, we may expect natural frequency differences to exist for some X-linked SNPs, along with deviations from HWE in females. Thus, it would be wrong to discard X SNPs due to an HWE violation, in case the violation can be explained as a natural result of sex-biased admixture.

In this work, we developed a likelihood ratio test and a χ^2 test for deviations from expected genotype frequencies on X, while permitting natural sex differences in frequency due to sex-biased admixture. This is achieved by taking into account the constraints imposed by Eqs. (1) and (2) on sex-specific frequencies across generations. We show by simulations that our tests have the expected size under the null, as well as power at least as high as existing tests for true deviations from the null (e.g., due to genotyping errors or inbreeding). Specifically, our tests reject the null substantially less often compared with existing tests when HWE is violated due to historical sex-biased admixture in otherwise randomly mating populations. Finally, we investigate the application of our tests to human data from the 1000 Genomes project.

Methods

We denote the number of males and females in the sample as n_m and n_f , respectively, and the two alleles as A and B. The numbers of male A and B carriers are denoted m_A and m_B . The numbers of females with genotypes AA, AB, and BB are denoted f_{AA} , f_{AB} , and f_{BB} . We denote by p_m and p_f the A allele frequencies in males and females, respectively.

We develop our likelihood ratio test based on the framework of You et al. (You et al. 2015). These authors have defined the *inbreeding coefficient* ρ to represent deviations from HWE. Using ρ , the expected genotype frequencies in females can be written as

$$p_{AA} = (1 - \rho)p_f^2 + \rho p_f = p_f^2 + \rho p_f(1 - p_f), \tag{5}$$

$$p_{AB} = 2p_f(1 - p_f)(1 - \rho), \tag{6}$$

$$\begin{aligned} p_{BB} &= (1 - \rho)(1 - p_f)^2 + \rho(1 - p_f) \\ &= (1 - p_f)^2 + \rho p_f(1 - p_f). \end{aligned} \tag{7}$$

The null hypothesis of no deviations from HWE and no frequency difference between males and females is $p_m = p_f = p$ and $\rho = 0$. We interpret here the parameter ρ more generally as a measure of the deviation from random mating in females, either positive or negative, and note that it is constrained to values that result in all genotype frequencies being in $[0, 1]$. The alternative hypothesis is $p_m \neq p_f$ or $\rho \neq 0$. Denote the parameters of the model as $\theta = (p_m, p_f, \rho)$. The likelihood of observing the data (genotype counts) is multinomial,

$$L(\theta) = \binom{n_f}{f_{AA}, f_{AB}, f_{BB}} p_{AA}^{f_{AA}} p_{AB}^{f_{AB}} p_{BB}^{f_{BB}} \times \binom{n_m}{m_A} p_m^{m_A} (1 - p_m)^{m_B}, \tag{8}$$

where p_{AA} , p_{AB} , and p_{BB} are given by Eqs. (5–7), respectively. You et al. have proposed an expectation–maximization algorithm to obtain the maximum-likelihood estimates (MLE) $\hat{\theta} = (\hat{p}_f, \hat{p}_m, \hat{\rho})$.

Under the null hypothesis, $p_m = p_f = p$ and $\rho = 0$, so $\theta_0 = (p, p, 0)$, and the likelihood reduces to

$$L(\theta_0) = \binom{n_f}{f_{AA}, f_{AB}, f_{BB}} \binom{n_m}{m_A} 2^{f_{AB}} p^{2f_{AA} + f_{AB} + m_A} (1 - p)^{2f_{BB} + f_{AB} + m_B}. \tag{9}$$

Here, the MLE is trivial, $\hat{\theta}_0 = (\hat{p}, \hat{p}, 0)$, where $\hat{p} = (2f_{AA} + f_{AB} + m_A) / (2n_f + n_m)$. The likelihood ratio (LR) statistic is

$$LR = 2 \ln \frac{L(\hat{\theta})}{L(\hat{\theta}_0)}. \tag{10}$$

The LR statistic is asymptotically distributed (under the null) as a χ^2 distribution with two degrees of freedom, leading to a test we call LRTP (likelihood ratio test for panmictic populations).

As explained above, the LRTP cannot accommodate “legitimate” frequency differences between the sexes due to sex-biased admixture. To address that, we reparametrize the model as follows. Instead of $\theta = (p_f, p_m, \rho)$, we write $\theta = (p_{f,g}, p_{m,g}, \rho)$, where $p_{f,g}$ and $p_{m,g}$ are the allele frequencies in females and males in the previous generation. With these parameters, the expected genotype frequencies in males in the current generation are

$$p_{A,m,c} = p_{f,g} \tag{11}$$

$$p_{B,m,c} = 1 - p_{f,g}. \tag{12}$$

This is analogous to Eq. (2), which is true because males receive X chromosomes only from females in the previous generation. Under random mating, the expected genotype frequencies in females in the current generation are

$$p_{AA,f,c} = p_{m,g}p_{f,g}, \tag{13}$$

$$p_{AB,f,c} = p_{m,g}(1 - p_{f,g}) + p_{f,g}(1 - p_{m,g}), \tag{14}$$

$$p_{BB,f,c} = (1 - p_{m,g})(1 - p_{f,g}). \tag{15}$$

The above expressions reflect the fact that females receive one X chromosome from males and one from females. To incorporate deviations from random mating, we use again the parameter ρ . Analogously to the case of panmictic populations, we write the expected genotype frequencies in females in the current generation as

$$p_{AA,f,c,\rho} = p_{m,g}p_{f,g} + \frac{\rho}{2}(p_{m,g}(1 - p_{f,g}) + p_{f,g}(1 - p_{m,g})), \tag{16}$$

$$p_{AB,f,c,\rho} = (p_{m,g}(1 - p_{f,g}) + p_{f,g}(1 - p_{m,g}))(1 - \rho), \tag{17}$$

$$p_{BB,f,c,\rho} = (1 - p_{m,g})(1 - p_{f,g}) + \frac{\rho}{2}(p_{m,g}(1 - p_{f,g}) + p_{f,g}(1 - p_{m,g})). \tag{18}$$

Note that the overall A allele frequency in females in the current generation is (for any ρ)

$$p_{f,c} = p_{AA,f,c} + \frac{p_{AB,f,c}}{2} = \frac{p_{m,g} + p_{f,g}}{2}, \tag{19}$$

as expected based on Eq. (1). Note that here too, ρ is constrained to values such that $p_{AA,f,c,\rho}$, $p_{AB,f,c,\rho}$, and $p_{BB,f,c,\rho}$

are all in [0, 1]. Our null hypothesis is that given the allele frequencies in the previous generation ($p_{f,g}$ and $p_{m,g}$), the genotypes of the current generation are determined by random mating, or $\rho = 0$. The alternative hypothesis is that there is a deviation from random mating, or $\rho \neq 0$. The likelihood of the data under the most general θ is

$$L(\theta) = \binom{n_f}{f_{AA}, f_{AB}, f_{BB}} p_{AA,f,c,\rho}^{f_{AA}} p_{AB,f,c,\rho}^{f_{AB}} p_{BB,f,c,\rho}^{f_{BB}} \times \binom{n_m}{m_A} p_{A,m,c}^{m_A} p_{B,m,c}^{m_B}, \tag{20}$$

where $p_{A,m,c}$, $p_{B,m,c}$, $p_{AA,f,c,\rho}$, $p_{AB,f,c,\rho}$, and $p_{BB,f,c,\rho}$ are defined by Eqs. (11), (12), (16–18), respectively. The MLE $\hat{\theta} = (\hat{p}_{f,g}, \hat{p}_{m,g}, \hat{\rho})$ is obtained by taking the derivatives of (the logarithm of) $L(\theta)$ and equating to zero. This results in a set of three equations, which are too tedious to reproduce here, and can be solved numerically to yield the MLE $\hat{\theta} = (\hat{p}_{f,g}, \hat{p}_{m,g}, \hat{\rho})$. In practice, we directly maximized the log-likelihood based on a grid search. (We discarded any parameter set $\hat{\theta}$ leading to allele frequencies in the current generation outside the range [0, 1] in Eqs. (16–18).)

In the case of random mating, $\rho = 0$, and thus the parameters are $\theta_0 = (p_{f,g}, p_{m,g}, 0)$. The likelihood is

$$L(\theta_0) = \binom{n_f}{f_{AA}, f_{AB}, f_{BB}} p_{AA,f,c}^{f_{AA}} p_{AB,f,c}^{f_{AB}} p_{BB,f,c}^{f_{BB}} \times \binom{n_m}{m_A} p_{A,m,c}^{m_A} p_{B,m,c}^{m_B}, \tag{21}$$

where $p_{AA,f,c}$, $p_{AB,f,c}$, and $p_{BB,f,c}$ are defined by Eqs. (13–15), respectively. Taking the derivatives of $\log L(\theta_0)$ with respect to $p_{f,g}$ and $p_{m,g}$ and equating to zero results in the following pair of equations,

$$\frac{f_{AA}}{\hat{p}_{m,g}} + \frac{f_{AB}(1 - 2\hat{p}_{f,g})}{(1 - \hat{p}_{m,g})(1 - \hat{p}_{f,g})} - \frac{f_{BB}}{1 - \hat{p}_{m,g}} = 0 \tag{22}$$

$$\frac{f_{AA}}{\hat{p}_{f,g}} + \frac{f_{AB}(1 - 2\hat{p}_{m,g})}{(1 - \hat{p}_{m,g})(1 - \hat{p}_{f,g})} - \frac{f_{BB}}{1 - \hat{p}_{f,g}} + \frac{m_A}{\hat{p}_{f,g}} - \frac{m_B}{1 - \hat{p}_{f,g}} = 0 \tag{23}$$

The solution of these equations yields the MLE under the null, $\hat{\theta}_0 = (\hat{p}_{f,g}, \hat{p}_{m,g}, 0)$. Here too, in practice we used a grid search to directly maximize the log-likelihood.

The likelihood ratio is then, as in Eq. (10),

$$LR = 2 \ln \frac{L(\hat{\theta})}{L(\hat{\theta}_0)}. \tag{24}$$

Under the null, the LR is asymptotically distributed as χ^2 with one degree of freedom, leading to a test we call LRTPA (for admixture).

Finally, we also propose a new χ^2 test, analogous to the χ^2 test proposed in (Graffelman and Weir 2016). Suppose

we have used Eqs. (22) and (23) to obtain the MLE $(\hat{p}_{f,g}, \hat{p}_{m,g})$. The expected values for the genotypes of males and females under the null ($\rho = 0$) are

$$\begin{aligned} & \text{Expected } f_{AA}, f_{AB}, f_{BB}, m_A, m_B \\ &= n_f \hat{p}_{m,g} \hat{p}_{f,g}, n_f (\hat{p}_{m,g} (1 - \hat{p}_{f,g}) + \hat{p}_{f,g} (1 - \hat{p}_{m,g})), \\ & \quad n_f (1 - \hat{p}_{m,g}) (1 - \hat{p}_{f,g}), n_m \hat{p}_{f,g}, n_m (1 - \hat{p}_{f,g}) \end{aligned} \quad (25)$$

Then, given the observed values of $f_{AA}, f_{AB}, f_{BB}, m_A, m_B$, a standard χ^2 statistic can be calculated, which would be asymptotically distributed as χ^2 with one degree of freedom. We call this test χ^2 -ML.

We also note that instead of the MLE $\hat{p}_{f,g}$ and $\hat{p}_{m,g}$, we could use a method of moments estimator, based on isolating $p_f(t)$ and $p_m(t)$ from Eqs. (1) and (2),

$$\hat{p}_{m,g} = 2p_f - p_m = \frac{2(2f_{AA} + f_{AB})}{2n_f} - \frac{m_A}{n_m}, \quad (26)$$

$$\hat{p}_{f,g} = p_m = \frac{m_A}{n_m}. \quad (27)$$

These estimates can then be substituted in Eq. (25), and a χ^2 statistic can be calculated. We call this test χ^2 -MM. In practice, we found that the χ^2 -MM did not appropriately control the type I error rate (Fig. 1), and we do not report further experiments with that test.

For comparison, we also considered the exact test of Graffelman and Weir for the X chromosome (Graffelman and Weir, 2016), which we denote GWET. The null hypothesis of that test includes both equality of allele

frequencies between the sexes, as well as HWE for the genotypes in females.

Results

Simulations

We carried out several simulations to examine the behavior of our tests as compared with existing tests for deviations from HWE on X. We considered scenarios under our tests' null hypothesis, as well as under a number of alternative hypotheses.

Our first set of simulations was designed to examine all tests under the null hypothesis of our tests, namely sex-biased admixture with random mating thereafter. We started with a population of 400 males and 400 females, and a single locus with an initial allele frequency of 80% in females and 30% in males. Given the allele frequencies in each generation, we calculated the expected genotype frequencies in the subsequent generation based on Eqs. (11–15). Then, the genotypes of 400 males and 400 females were drawn based on multinomial distributions having these expected frequencies. We repeated the process up to ten generations after admixture, and repeated the simulation 1000 times.

In Fig. 1a, we show the proportion of rejections (type I error rate, or size) when running five tests on the above genotype counts: the LRTP test of You et al. (You et al. 2015) and the GWET (exact test) of Graffelman and Weir

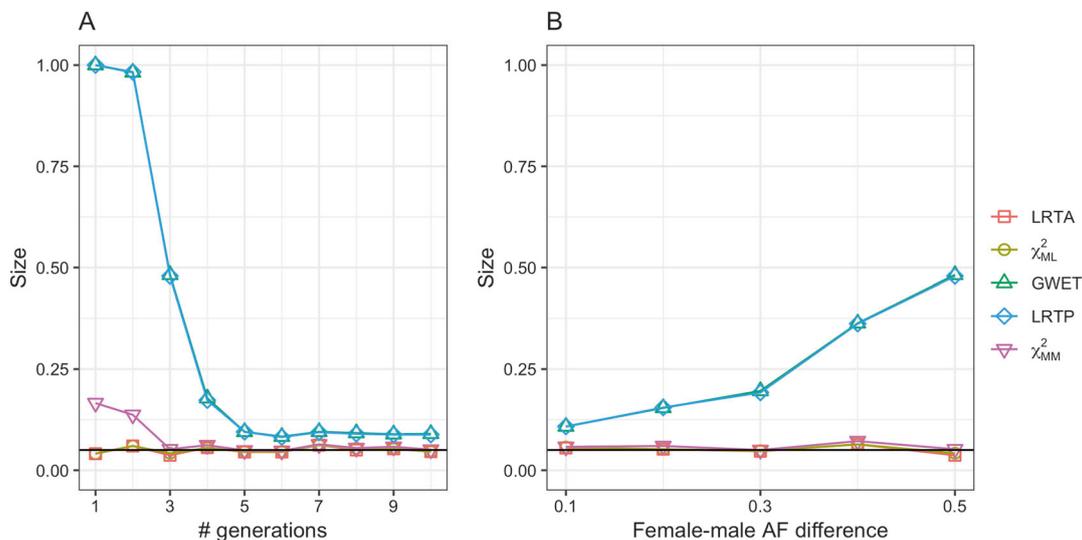


Fig. 1 The proportion of rejections (type I error rate, or size) under random mating in a sex-biased admixed population. We compared the LRTP test (You et al. 2015), the GWET exact test (Graffelman and Weir, 2016), and the LRTA, χ^2 -ML, and χ^2 -MM tests developed in this paper. Our significance level was $\alpha = 0.05$ (horizontal black line). **a**

The size of the tests versus the number of generations since admixture. The initial allele frequencies were 80% in females and 30% in males. **b** The size of the tests versus the initial allele frequency (AF) difference between males and females. The number of generations since admixture was three

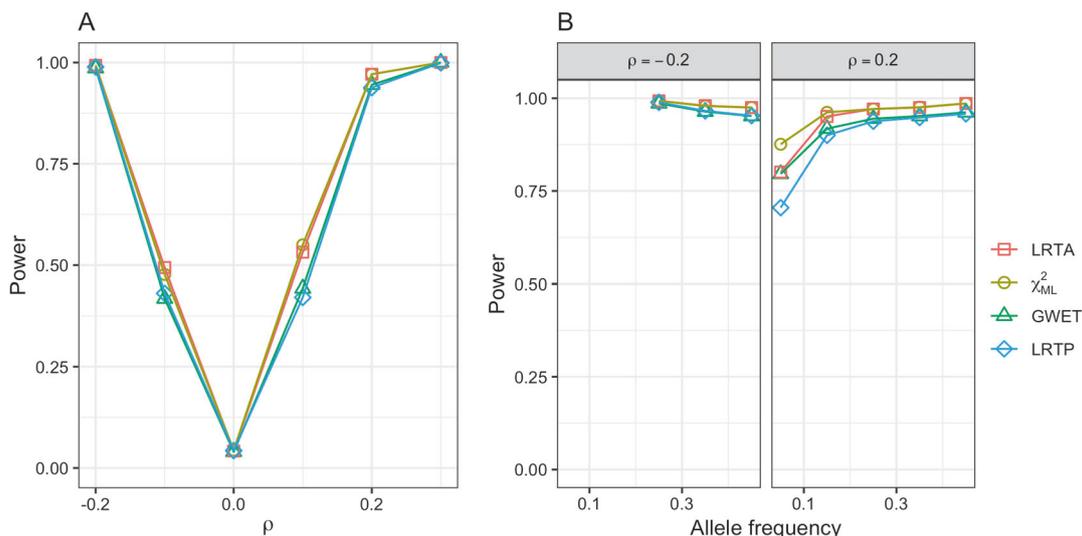


Fig. 2 The proportion of rejections (power) of the various tests under non-random mating (without admixture). **a** Different values of the inbreeding coefficient ρ , for a fixed allele frequency (0.8, equal between males and females). **b, c** Different values of the allele

frequency for a fixed inbreeding coefficient ρ (b: -0.2 ; c: 0.2). For $\rho = -0.2$, some values of the allele frequencies are missing as they would lead to genotype frequencies outside $[0, 1]$

(Graffelman and Weir 2016), both of which test for departures from either HWE in females or equality of allele frequencies between males and females; and the LRTA, χ^2 -ML, and χ^2 -MM tests we have developed here for sex-biased admixed populations (see the Methods section). Our LRTA test and the χ^2 -ML test had an appropriate type I error rate (equal or close to the significance level $\alpha = 0.05$), which is expected, because we simulated random mating post admixture. In contrast, the LRTP and GWET tests had much higher proportions of rejections, as expected due to the frequency differences between the sexes, which these tests are designed to detect. For these tests, the type I error rate decreased to a value close to the appropriate value under the null (0.05) after about ≈ 5 – 6 generations post admixture. The χ^2 -MM test did not control the type I error rate as well as the LRTA and χ^2 -ML tests, possibly because the parameters (allele frequencies in the preceding generation) were not accurately estimated. We thus do not further consider this test.

In Fig. 1b, we plot the type 1 error rate versus the initial allele frequency difference between the sexes, for a sample taken three generations post admixture. Again, our LRTA and χ^2 -ML tests control the type 1 error rate at its appropriate value (0.05), while the LRTP and GWET tests have a higher proportion of rejections, which is growing, as expected, with the initial allele frequency difference between the sexes.

Our second simulation was designed to examine the power of the various tests under the alternative hypothesis of non-random mating. We considered one locus with an allele frequency of 80% in both males and females. We then calculated the expected genotype frequencies under one

generation of mating, but this time with an inbreeding coefficient ρ in the range $[-0.2, 0.3]$, and simulated genotype frequencies in 400 females and 400 males based on the multinomial distribution with probabilities defined by Eqs. (11), (12), (16–18). This simulation did not include sex-biased admixture, as the goal was to evaluate the power of our test under non-random mating, regardless of a history of admixture. We present the power of the various tests (at the 0.05 significance level and over 1000 repeats) in Fig. 2a. The power of the χ^2 -ML test is always the highest, followed closely by the LRTA test. The power of the LRTP and GWET tests is slightly lower. In Fig. 2b and 2c, we considered the cases $\rho = -0.2$ (panel b) and $\rho = 0.2$ (panel c), for a varying allele frequency (equal between the sexes). The results are similar, again with our LRTA and χ^2 -ML tests achieving the highest power, followed closely by the existing LRTP and GWET tests.

Our third set of simulations was designed to validate that the LRTA and χ^2 -ML tests are powerful also under sex-biased admixture. We used the same approach as in our first simulation (Fig. 1), i.e., sex-biased admixture followed by random mating, except that after one generation, non-random mating was assumed with an inbreeding coefficient in the range $[-0.2, 0.3]$. The initial male allele frequency was 0.3, and the initial female allele frequency was 0.4 or 0.8. We plot in Fig. 3 the power of the various tests at the 0.05 significance level and over 50 repeats. The power of our LRTA and χ^2 -ML tests is almost unaffected by the historical admixture event (compare Fig. 2a). When the initial allele frequency difference between the sexes is large (panel b) and for small $|\rho|$, the LRTP and GWET tests have a much higher power than our tests, as expected. This is

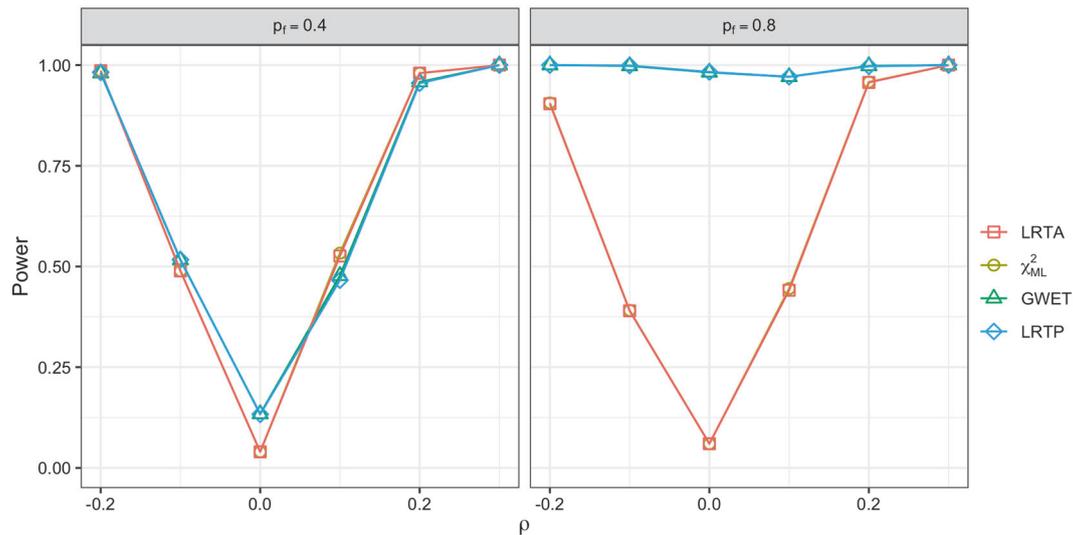


Fig. 3 The proportion of rejections (power) of the various tests under non-random mating and sex-biased admixture. The power is plotted for different values of the inbreeding coefficient ρ , for a given initial male allele frequency of 0.3. The initial female allele frequency is 0.4 in (a) and 0.8 in (b). The power of our LRTA and χ^2 -ML tests is similar to that reported in Fig. 2 for the case without admixture, with a slight drop in power for a large initial allele frequency difference

between the sexes (panel b). In that regime, the power of the LRTP and GWET tests is high for all values of ρ , because these tests are identifying frequency differences between the sexes and deviations from HWE in females. However, as these deviations are generated (for small $|\rho|$) by sex-biased admixture, they are correctly ignored by our methods, which as expected have higher power with increasing $|\rho|$

because these tests are identifying (as they are expected to) the allele frequency differences between the sexes and the deviations from HWE in females that were generated by the sex-biased admixture. Our tests reject the null only for larger values of $|\rho|$, which is the correct and expected behavior, given that our tests do not reject the null when the deviations from HWE or equal allele frequency are due to sex-biased admixture.

In our last set of simulations, we sought to determine whether our tests can reject the null under deviations other than nonzero inbreeding coefficients. To this end, we simulated genotype counts for 100 males and 100 females under an allele frequency of 0.5, equal between the sexes. We then flipped each female AA genotype to AB with probability q , and compared the power of the various tests at increasing levels of the genotype flipping probability q . The results (Fig. 4a) demonstrate that our LRTA and χ^2 -ML tests are powered to detect these deviations, although less so than the existing tests. In contrast, when simulating 200 males and 200 females (Fig. 4b), the power of our tests was comparable with the power of the existing tests.

Empirical data analysis

To examine our tests on real data, we applied them to genomes from the 1000 Genomes project (1000 Genomes Project Consortium et al. 2015) (1kG). We selected American populations with a history of sex-biased, cross-continental recent admixture. While admixture in these

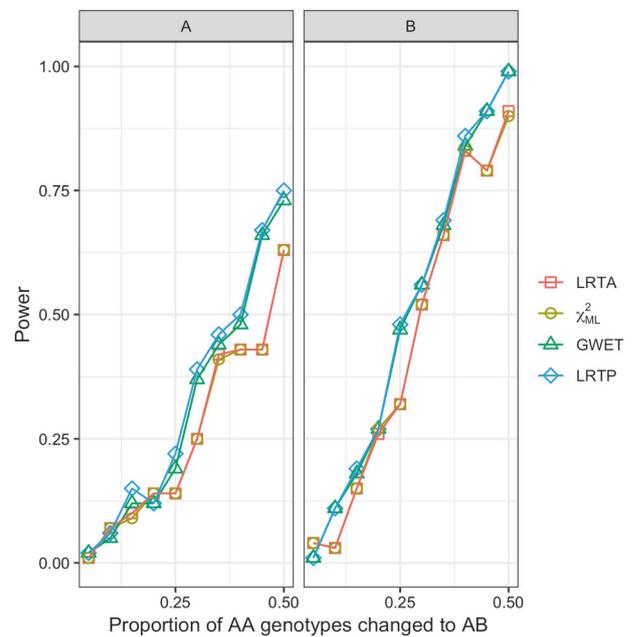


Fig. 4 The power of the various tests to detect genotyping errors. **a** We simulated genotype counts for 100 males and 100 females under allele frequency 0.5 in both sexes. We then randomly flipped each female AA genotype to AB with probability q (the x-axis), and computed the power of the various tests over 100 repeats. Panel (b) is the same as (a), for 200 males and 200 females

populations has mostly ended 5–10 generations ago (e.g., Baharian et al. 2016; Gravel 2012; Gravel et al. 2013; Moreno-Estrada et al. 2013), some SNPs may have not yet reached equilibrium, or were affected by more recent minor

Table 1 The proportion of rejected SNPs (at $\alpha = 0.05$) under the LRTP, GWET, LRTA, and χ^2 -ML tests in 1 kG American populations

Population	#SNPs	LRTP	GWET	LRTA (this paper)	χ^2 -ML (this paper)
ACB (African Caribbeans in Barbados)	779,444	0.031	0.030	0.025	0.036
ASW (Americans of African Ancestry in SW USA)	692,300	0.028	0.027	0.024	0.036
CLM (Colombians from Medellin, Colombia)	549,946	0.037	0.037	0.032	0.045
MXL (Mexican Ancestry from Los Angeles USA)	415,094	0.027	0.026	0.028	0.035
PEL (Peruvians from Lima, Peru)	449,003	0.024	0.022	0.022	0.026
PUR (Puerto Ricans from Puerto Rico)	617,842	0.034	0.033	0.029	0.039

For each population, the lowest proportion of rejected SNPs is indicated in bold

Table 2 Investigating HWE for two X chromosome SNPs

SNP	AA	AB	BB	A	B	p_f	p_m	HWE	LTRA	LRTP	$p_{f,g}$	$p_{m,g}$	E-AA	E-AB	E-BB	E-A	E-B
1	27	22	0	42	5	0.78	0.89	0.04	0.06	0.008	0.92	0.6	27.0	20.4	1.6	43.2	3.8
2	17	30	2	38	9	0.65	0.81	0.01	0.12	0.005	0.85	0.42	17.5	27.2	4.3	40.0	7.1

For each SNP, we provide genotype counts in females (AA, AB, and BB) and males (A, B) in the ACB 1000 Genomes population, allele frequencies in females and males (p_f and p_m , respectively), the P -value for a deviation from HWE in females (χ^2 test), and the P -values from our LRTA test and the LRTP test. The rightmost seven columns provide information on the null hypothesis of our LRTA test. Specifically, $p_{f,g}$ and $p_{m,g}$ are the maximum-likelihood estimates of the female and male allele frequencies in the previous generation, respectively. The rightmost five columns are the expected genotype counts in females (E-AA, E-AB, and E-BB) and males (E-A, E-B) under random mating, given the allele frequencies in the previous generation (compare with the actual genotype counts in columns 2–6)

gene flow events. Our goal was to determine whether our tests reject fewer SNPs due to deviations from expected genotype frequencies. Indeed (Table 1), our LTRA test rejected the lowest proportion of SNPs in nearly all populations. The LRTP and GWET test showed a similar higher proportion of rejected SNPs, while our χ^2 -ML test rejected the largest proportion of SNPs in all populations.

To demonstrate the underlying reasons for the lower rejection rates, we considered the following two examples from the ACB population: SNP 1 at coordinate 116399355 (hg19), and SNP 2 at coordinate 78068904. The genotype counts, allele frequencies, and P -values are presented in Table 2. For both SNPs, a substantial allele frequency difference exists between the sexes, along with deviation from HWE in females, and hence the null is rejected by LRTP (as well as by GWET; not shown). In contrast, the null is rejected neither by our LRTA nor by the χ^2 -ML test (not shown). To see why, consider the rightmost seven columns of the table, where the estimated allele frequencies in the previous generation are shown, along with the expected genotype counts in the current generation (assuming random mating). In both cases, the expected genotype counts are similar to the observed ones, thus not providing a reason to reject the null hypothesis of random mating. However, we note that in these two examples, the implied allele frequency difference between the sexes in the previous generation is relatively high, at 0.32 and 0.43 in SNPs 1 and 2, respectively, and the plausibility of such a large natural difference should be carefully considered by users of the tests. A similar picture was observed in other SNPs, where

often, SNPs were rejected by the existing tests, but not by our LRTA or χ^2 -ML tests, on the basis of a very large allele frequency difference between the sexes in the previous generation (see the Discussion section).

Discussion

In this paper, we proposed new tests for deviations from expected genotype frequencies on the X chromosome for sex-biased admixed populations. The X chromosome is unique in that allele frequencies do not reach equilibrium within one generation after perturbation, even when the population is otherwise randomly mating and all genotypes are observed without error. Thus, the X chromosome requires a specialized test, even beyond accounting for the different ploidies of the sexes. Here, we proposed new likelihood ratio and χ^2 tests to address this gap. We showed that our tests have the expected size (type I error rate) under sex-biased admixture followed by random mating, whereas other tests have high error rates, in particular when admixture was very recent (Fig. 1). In addition, our tests had equal or higher power compared with the other tests to detect nonzero inbreeding coefficients (Fig. 2). Our tests were powered, although less than existing tests, to detect nonzero inbreeding coefficients under strong sex-biased admixture (Fig. 3), as well as errors introduced by directionally flipping one of the genotypes for small samples (Fig. 4).

We demonstrated that our tests rejected fewer X chromosome SNPs in American 1000 Genomes populations.

However, there are a number of important caveats. One issue is that admixture in these populations has likely ended already 5–10 generations ago. Considering Eqs. (3) and (4) (for the case when the allele is present in all female founders but in none of the male founders), the difference in allele frequency between males and females is just $\approx 3\%$ after five generations. This difference is already smaller than the standard deviation of the difference in allele frequency between 200 males and 200 females ($\approx 4.7\%$; assuming true frequency $2/3$ in both sexes, which is the equilibrium solution of Eqs. (3) and (4)). Thus, it cannot be established which SNPs have different allele frequency between the sexes due to sex-biased admixture, as opposed to other causes such as sampling noise or subtle ethnicity differences between males and females in the sample.

The second (and related) issue is that for most SNPs that were rejected by the existing methods but not by the tests presented here, the implied allele frequency difference between males and females in the previous generation was very high. Large differences in allele frequencies in the previous generation are permitted by our tests, as frequencies could indeed widely differ between males and females ≈ 1 – 2 generations post admixture. However, in human populations, sex-biased admixture was not as recent, and thus, the relevance of the test to human genetic data remains at this point unclear. Another limitation is the relatively low power of our test to detect genotype “flipping” for small sample sizes.

Finally, we also note that when applying our tests to genomic data, care must also be taken if the downstream application is an association test that relies on HWE, such as the alleles test (Laird and Lange 2011), because SNPs may pass our tests even if females are not in HWE.

Our tests are available as an R package called HWadmiX, at <https://github.com/dbackenroth/HWadmiX>. The running time of computing both the LRTA and χ^2 -ML P -values was on average ≈ 1.1 s per SNP for the 49 females and 47 males of the ACB population, on a computer with a 2.5 GHz Intel Core i7 CPU and 16 GB RAM. To reduce the running time when testing all X chromosome 1000 Genomes Project variants, we pre-calculated the P -values for all combinations of observed genotype counts. The number of combinations to pre-calculate is at most $O(n_f^2 n_m)$, which is smaller than the number of variants for the 1000 Genomes sample sizes.

Avenues for future research include replacing the grid search for the maximum-likelihood estimate by a faster method, as well as the development of tests for multiple alleles or Bayesian tests. Specifically, it would be of interest to impose a prior distribution on the allele frequency difference between females and males in the previous generation, which could make the test more appropriate to

human populations, where sex-biased admixture occurred centuries ago. Finally, exact tests for HWE are in widespread use (Purcell et al. 2007) and were also developed for HWE on X (Graffelman and Weir 2016). Developing an exact test for sex-biased admixed populations will be of interest, as these tests better control the type 1 error rate (Wigginton et al. 2005). However, it does not seem that the approach of (Graffelman and Weir 2016) can be easily extended to admixed populations.

Acknowledgements We thank Alon Keinan for discussions. SC thanks the German–Israeli Foundation for Scientific Research and Development (GIF) grant I-2489-407.6/2017 and the Israel Science Foundation (ISF) grant 407/17.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM et al. (2015) A global reference for human genetic variation. *Nature* 526:68–74
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564–1573
- Ayres KL, Balding DJ (1998) Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Hered (Edinb)* 80(Pt 6):769–777
- Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ et al. (2016) The great migration and African-American genomic diversity. *PLoS Genet* 12:e1006059
- Bonnen PE, Lowe JK, Altshuler DM, Breslow JL, Stoffel M, Friedman JM et al. (2010) European admixture on the Micronesian island of Kosrae: lessons from complete genetic information. *Eur J Hum Genet* 18:309–316
- Bourgain C, Abney M, Schneider D, Ober C, McPeck MS (2004) Testing for Hardy-Weinberg equilibrium in samples with related individuals. *Genetics* 168:2349–2361
- Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A et al. (2010) Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA* 107(Suppl 2):8954–8961
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562:203–209
- Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ et al. (2014) Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS ONE* 9:e113684
- Emigh TH (1980) A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* 36:627–642
- Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z et al. (2015) XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J Hered* 106:666–671

- Graffelman J (2015) Exploring diallelic genetic markers: the Hardy-Weinberg Package. *J Stat Softw* 64:1
- Graffelman J, Jain D, Weir B (2017) A genome-wide study of Hardy-Weinberg equilibrium with next generation sequence data. *Hum Genet* 136:727–741
- Graffelman J, Sanchez M, Cook S, Moreno V (2013) Statistical inference for Hardy-Weinberg proportions in the presence of missing genotype information. *PLoS ONE* 8:e83316
- Graffelman J, Weir BS (2016) Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Hered (Edinb)* 116:558–568
- Graffelman J, Weir BS (2018) Multi-allelic exact tests for Hardy-Weinberg equilibrium that account for gender. *Mol Ecol Resour* 18:461–473
- Gravel S (2012) Population genetics models of local ancestry. *Genetics* 191:607–619
- Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL et al. (2013) Reconstructing native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9:e1004023
- Hao W, Storey JD (2017) Extending tests of Hardy-Weinberg equilibrium to structured populations. *bioRxiv*. <https://doi.org/10.1101/240804>
- Hernandez JL, Weir BS (1989) A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* 45:53–70
- Jagadeesan A, Gunnarsdottir ED, Ebenesersdottir SS, Guethmundsdottir VB, Thordardottir EL, Einarsdottir MS et al. (2018) Reconstructing an African haploid genome from the 18th century. *Nat Genet* 50:199–205
- Jennings HS (1916) The numerical results of diverse systems of breeding. *Genetics* 1:53–89
- Khrantsova EA, Davis LK, Stranger BE (2019) The role of sex in the genomics of human complex traits. *Nat Rev Genet* 20:173–190
- Kim SK, Gignoux CR, Wall JD, Lum-Jones A, Wang H, Haiman CA et al. (2012) Population genetic structure and origins of native Hawaiians in the multiethnic cohort study. *PLoS ONE* 7:e47881
- Kudelka MR, Hinrichs BH, Darby T, Moreno CS, Nishio H, Cutler CE et al. (2016) Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. *Proc Natl Acad Sci USA* 113:14787–14792
- Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z, Knowles DA et al. (2016) Impact of the X Chromosome and sex on regulatory variation. *Genome Res* 26:768–777
- Laird NM, Lange C (2011) The fundamentals of modern statistical genetics. Springer-Verlag, New York
- Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhargale T et al. (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* 34:591–602
- Levene H (1949) On a matching problem arising in genetics. *Ann Math Stat* 20:91
- Li YR, Li J, Zhao SD, Bradfield JP, Mentch FD, Maggadottir SM et al. (2015) Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 21:1018–1027
- Lie BA, Dupuy BM, Spurkland A, Fernandez-Vina MA, Hagelberg E, Thorsby E (2007) Molecular genetic studies of natives on Easter Island: evidence of an early European and Amerindian contribution to the Polynesian gene pool. *Tissue Antigens* 69:10–18
- Lind JM, Hutcheson-Dilks HB, Williams SM, Moore JH, Essex M, Ruiz-Pesini E et al. (2007) Elevated male European and female African contributions to the genomes of African American individuals. *Hum Genet* 120:713–722
- Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD et al. (2016) A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun* 7:12522
- Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR et al. (2013) Reconstructing the population genetic history of the Caribbean. *PLoS Genet* 9:e1003925
- Puig X, Ginebra J, Graffelman J (2017) A Bayesian test for Hardy-Weinberg equilibrium of biallelic X-chromosomal markers. *Hered (Edinb)* 119:226–236
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Rohlf R, Weir BS (2008) Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics* 180:1609–1616
- Rosenberg NA (2016) Admixture models and the breeding systems of H. S. Jennings: A GENETICS Connection. *Genetics* 202:9–13
- Scelsi MA, Khan RR, Lorenzi M, Christopher L, Greicius MD, Schott JM et al. (2018) Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci. *Brain* 141:2167–2180
- Shriner D (2011) Approximate and exact tests of Hardy-Weinberg equilibrium using uncertain genotypes. *Genet Epidemiol* 35:632–637
- Traglia M, Bseiso D, Gusev A, Adviento B, Park DS, Mefford JA et al. (2017) Genetic mechanisms leading to sex differences across common diseases and anthropometric traits. *Genetics* 205:979–992
- Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT et al. (2011) Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet* 1–18. <https://doi.org/10.1002/0471142905.hg0119s68>
- Wakefield J (2010) Bayesian methods for examining Hardy-Weinberg equilibrium. *Biometrics* 66:257–265
- Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76:887–893
- Yap CX, Sidorenko J, Wu Y, Kemper KE, Yang J, Wray NR et al. (2018) Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nat Commun* 9:5407
- You XP, Zou QL, Li JL, Zhou JY (2015) Likelihood ratio test for excess homozygosity at marker loci on X chromosome. *PLoS ONE* 10:e0145032
- Yu C, Zhang S, Zhou C, Sile S (2009) A likelihood ratio test of population Hardy-Weinberg equilibrium for case-control studies. *Genet Epidemiol* 33:275–280
- Zheng G, Joo J, Zhang C, Geller NL (2007) Testing association for markers on the X chromosome. *Genet Epidemiol* 31:834–843