Published in final edited form as:

Genet Epidemiol. 2005 December ; 29(4): 377–388. doi:10.1002/gepi.20093.

Estimating the additive genetic effect of the X chromosome

Jack W. Kent Jr.^{*}, Thomas D. Dyer, and John Blangero

Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas
USA

Abstract

We propose a method for efficient estimation of the additive genetic effect of the X chromosome with explicit modeling of eutherian-type dosage compensation. The theoretical derivation of the variance-components model for X-linked loci is reviewed in detail. We develop a model of dosage compensation that allows for both incomplete and heterogeneous lyonization, the existence of which is suggested by recent expression studies. Modeling this relationship, especially in the limit cases of complete or absent compensation, allows estimation of the X effect as a single parameter for ease of comparison to other sources of variance. We present simulation studies to estimate the power and computational efficiency of our proposed method.

Keywords

sex linkage; X-chromosome; dosage compensation; variance components; quantitative trait;
statistical genetics

INTRODUCTION

Because segregation of the sex chromosomes is directly associated with a readily observed phenotype (sex), X-linked genes were among the first to be localized during the first century of modern genetics. In spite of this, analysis of the X chromosome has not been fully integrated into variance-components-based genetic software, due to the difficulty of analyzing male/female asymmetries in the transmission and dosage of X-linked genes. Because of these analytical complexities, investigators may give lower priority to genotyping X-chromosomal markers in large studies, thereby compounding the obstacles to gene discovery in this part of the genome.

This situation is beginning to change. Ekstrøm [2004] proposed an algorithm for multipoint linkage analysis on the X chromosome. In part to avoid making assumptions about the effects of gene dosage, Ekstrøm's method estimates separate variance components for male-male, female-female, and male-female relative pairs, with separate identity-by-descent (IBD) matrices for each class of paired individuals. This pattern is applied both to the linkage parameters at a given locus and to the residual additive genetic variance due to other X-linked loci.

We propose a modified method that explicitly parameterizes the effect of dosage compensation in terms of *lyonization coefficients* (named for Mary Lyon, an early proponent of the X-inactivation hypothesis [Lyon 1961]). Our derivation of these coefficients is general enough to allow for incomplete inactivation, suggested by recent

^{*}Corresponding author: Jack W. Kent Jr., Ph.D., Dept. of Genetics, SFBR, PO Box 760549, San Antonio, Texas 78245-0549, jkent@darwin.sfbr.org, (210) 258-9570.

expression studies [Carrel & Willard 2005]. However, we also consider a more parsimonious approach that builds on Bulmer's [1985] observations on the linear relationship between male and female variances in the complete presence or absence of dosage compensation, as occurs in the majority of X-linked loci in eutherian females. Bulmer's simple alternate models for these cases allow us to estimate the additive genetic effect of the X chromosome as a single parameter, drawing information from all classes of paired individuals in the pedigree, in a way that should be appropriate for most mammalian systems (and many other organisms as well). We anticipate that any loss of generality in our proposed model may be compensated by the efficiency of estimating a minimal number of new parameters. We report simulation studies designed to estimate the power of our proposed method, including the cost in power when we have misjudged the presence or absence of dosage compensation in a particular case.

Our emphasis in this paper is on estimation of the aggregate additive genetic effect of X-linked loci, for three reasons. (1) The estimation of this effect is necessary for complete specification of the linkage model, and the derivation of expected allele-sharing will inform the calculation of locus-specific IBDs. (2) Inclusion of the X effect may improve model specification even when the focus is on autosomal linkages. (3) Finally and importantly, we wish to give the investigator a rapid estimate of the magnitude of the X-chromosomal contribution to a phenotype of interest, using pedigree data alone, to determine the priority of a more detailed analysis of this chromosome.

THEORETICAL BACKGROUND

The additive effect of X-linked genes is a function of two distinct processes: the transmission of alleles from parent to offspring, and the expression of those alleles. For variance-component analysis in general pedigrees, the decomposition of the phenotypic covariance between relatives should include an X-chromosome-specific variance term (in general, separate variances for each sex) weighted by factors representing the effects of allele-sharing and dosage compensation. Briefly, these factors are:

- a. an X-linkage analog of the autosomal matrix Φ of probabilities that an allele drawn from a locus in one individual is identical by descent (IBD) with that drawn from another individual (or drawn with replacement from the same individual). Following the notation in [Ekström 2004], we call this matrix Ψ . The elements of Ψ can be calculated recursively from pedigree data, as by the algorithm in MINX ("MERLIN in X") [Abecasis 2005];
- b. IBD weighting coefficients that reflect the unequal copy number of the X in males and females;
- c. weighting coefficients, which we call *lyonization coefficients*, that account for the dosage-compensation state of the locus or loci under analysis.

The derivation of Ψ and the IBD weights is well-established in the literature [Bulmer 1985, Grossman & Eisen 1989, Cordell et al. 1995, Lynch & Walsh 1998], although there is some disagreement on the weighting for male-female pairs (see below). Because of this disagreement (and some obscurity!) in the literature, we review the derivation below.

The effects of dosage compensation usually have been modeled assuming either complete inactivation (lyonization) of one copy of the X in females or the complete absence of inactivation [Bulmer 1985, Pérez-Enciso et al. 2002]. This assumption is reasonable for most loci in eutherian females, but recent evidence suggests that inactivation may be incomplete at some loci [Carrel & Willard 2005]. Based on expression profiling in human fibroblast cell lines, these authors estimate that ~75% of X-linked loci are liable to

inactivation, while ~15% escape inactivation (although inactivation may be incomplete). Unexpectedly, ~10% exhibit rates of inactivation that vary widely among individual females. Consequently, we present a model of dosage compensation that is general enough to allow for incomplete inactivation (below), and can be generalized further to include heterogeneous inactivation (see Appendix). Explicit modeling of dosage compensation establishes a simple mathematical relationship between the male and female variances due to the X, allowing expression of the additive genetic effect of the X in all individuals in terms of the variance in one sex [Bulmer 1985], i.e., as a single additional parameter. We present simulation studies that test the power, accuracy, and computational efficiency of this strategy.

ALLELE SHARING

Allele sharing: female-female pairs—Both individuals i, j are diploid for the X, so the derivation is the same as the autosomal case. The following derivation mirrors that for autosomal loci as presented in a standard text [Lynch & Walsh 1998, p. 143]. For a given locus q , and assuming no dosage compensation,

$$\text{cov}_q(i, j) = E[(\alpha_{q1,i} + \alpha_{q2,i} - \mu_i)(\alpha_{q1,j} + \alpha_{q2,j} - \mu_j)]$$

where each α is the phenotypic deviation due to the additive effect of an allele at the locus. Because $\mu_i = \mu_j = 0$ for deviations from the phenotypic mean,

$$\text{cov}_q(i, j) = E[(\alpha_{q1,i} + \alpha_{q2,i})(\alpha_{q1,j} + \alpha_{q2,j})] = \sum_k \sum_m E[\alpha_{qk,i} \alpha_{qm,j}], k \in \{1, 2\}; m \in \{1, 2\}.$$

Each of the four elements in the summation is $E(\alpha_q^2)$ with probability ψ_{ij} if the alleles are IBD (where ψ_{ij} is the X-linked analog of ϕ_{ij}), or $[E(\alpha_q)]^2 = 0$ otherwise. Therefore,

$$\text{cov}_q(i, j) = 4\psi_{ij}E(\alpha_q^2).$$

Summing over loci, $E(\alpha_q^2) = \sigma_{AX,ff}^2/2$, or one-half the additive genetic variance due to X-linked loci in female-female pairs (each α_i accounts for the per-locus contribution of one of the two copies of the X chromosome in each individual). Thus,

$$\text{cov}_X(i, j) = 2\Psi_{ff}\sigma_{AX,ff}^2 \quad (1)$$

As in the autosomal case, the elements of Ψ_{ff} are $P_1/4 + P_2/2$, where P_1 and P_2 are the respective probabilities of sharing one or two alleles IBD.

Allele sharing: male-male pairs—Both individuals are haploid for the X chromosome. For a given locus q ,

$$\begin{aligned} \text{cov}_q(i, j) &= E[(\alpha_{q,i})(\alpha_{q,j})] \\ &= \psi_{ij}E(\alpha_q^2). \end{aligned}$$

Summing over loci, $E(\alpha_q^2) = \sigma_{AX,mm}^2$ (only one chromosome provides all of the X-linked additive genetic variance in each individual), and

$$\text{cov}_X(i, j) = \Psi_{mm} \sigma_{AX,mm}^2 \quad (2)$$

The elements of Ψ_{mm} are simply P_1 .

Allele sharing: male-female pairs—Note that the male is haploid for the X while the female is diploid. For a given locus q ,

$$\begin{aligned} \text{cov}_q(i, j) &= E[(\alpha_{q,i})(\alpha_{q1,j} + \alpha_{q2,j})] = \sum_k E[\alpha_{q,i} \alpha_{qk,j}], k \in \{1, 2\} \\ &= 2\psi_{ij} E(\alpha_q^2). \end{aligned}$$

Summing over loci, $E(\alpha_q^2) = \sigma_{AX,ff}^2/2$ for the female relative *but* $\sigma_{AX,mm}^2$ for the male relative. Therefore the covariance becomes,

$$\text{cov}_X(i, j) = 2\Psi_{mf} [\sigma_{AX,mm}^2 (\sigma_{AX,ff}^2/2)]^{0.5} \rho_{Xg,mf} \quad (3a)$$

$$= (\sqrt{2}) \Psi_{mf} \sigma_{AX,mm} \sigma_{AX,ff} \rho_{Xg,mf} \quad (3b)$$

$\rho_{Xg,mf}$, the correlation of the values of specific genotypes in males and females, can differ from 1 due to X-linked-genotype $\times$ sex interactions (or heterogeneous inactivation: see Appendix). The elements of Ψ_{mf} are $P_1/2$. If $\rho_{Xg,mf} = 1$, Equation 3b can be written as,

$$\text{cov}_X(i, j) = (1/\sqrt{2}) P_1 \sigma_{AX,mm} \sigma_{AX,ff} \quad (3c)$$

Equation 3c is the “general case” for male-female covariance for X-linked loci as presented (without derivation) by Bulmer [1985].

Equation 3b has a coefficient of $\sqrt{2}$ because, in the absence of dosage compensation, an X-linked allele accounts for a different proportion of the trait variance in males and females. This is in agreement with some presentations [Bulmer 1985, Grossman & Eisen 1989, Cordell et al. 1995] but not others [Lynch & Walsh 1998, Ekström 2004], where a coefficient of 2 is used. Our notation also differs from that of [Ekström 2004] by explicitly stating the covariance term as $\sigma_{AX,mm} \sigma_{AX,ff} \rho_{Xg,mf}$.

DOSAGE

Now we must deal with possible differences in the expression of X-linked alleles due to their different copy number in males and females. Many species have mechanisms for equalizing the dosage of X-linked genes. In some invertebrates, for example, both copies of the X may be downregulated in homogametic individuals (*C. elegans*), or the single X in the heterogametic sex may be upregulated (*Drosophila* spp.) [Amrein 2000]. In female eutherian mammals, a majority of loci on one copy of the X chromosome are downregulated through X-inactivation (lyonization). As presently understood, this inactivation is random with respect to the maternal or paternal copy of the X in different cell lineages, and also random with respect to the timing of inactivation during development; thus a female is a mosaic of patches of cells expressing one or the other X-haplotype.

A general model of eutherian X-inactivation (which can also be applied to any system in which compensation is mosaic and random with respect to the parental origin of the X-homologs) is as follows: Let there be two alleles for an X-linked locus q , with additive effects scaled such that allele A1 contributes 1 to a trait, while A2 contributes 0. Let p = frequency of A1. In males, the single allele is expressed with probability 1. In females, one allele in each cell lineage is fully expressed, while the other is expressed with probability λ_q (or, equivalently, expressed at a fraction λ_q of its maximum possible value: $0 \leq \lambda_q \leq 1$). If λ_q is a constant, and because each female is mosaic with each allele equally liable to inactivation in different cell lineages, the mean expressivity due to lyonization of each genotype at q is $(1 + \lambda_q)/2$. The resulting phenotypes and their means and variances are shown in Table 1. The Appendix addresses the more complex case where λ_q is a random variable (i.e., where inactivation is heterogeneous).

Classical models of dosage compensation or its absence, such as those of Bulmer [1985], are equivalent to the limit cases of our model where λ_q is either 0 or 1. As these cases are relatively simple (and good alternate hypotheses for most X-linked loci), we have employed them in the simulations that follow. The male and female X-linked additive genetic variances are linearly related ([Bulmer 1985] and Table 1). This allows us to express the additive effect of the X in the population in terms of its effect in one or the other sex – say, for example, the female (the following discussion assumes that females are the homogametic sex):

$$\text{cov}_X(x, y) = \Psi \odot \mathbf{X} \sigma_{AX,ff}^2, \quad (4)$$

where $\odot$ represents the Hadamard matrix product operator (yielding the elementwise product $\psi_{i,j} \times X_{i,j}$ for each pair of individuals i and j). Thus, the structuring element is the Hadamard product of a combined Ψ matrix and a coefficient matrix $\mathbf{X}$, each element of which is itself a product of coefficients $L_i L_j$. Each $L_{i,j}$ is the coefficient of Ψ required by allele-sharing for two individuals i and j , appropriate to their respective sexes (these coefficients were derived in the preceding section). L_i and L_j are *lyonization coefficients* that allow the single variance parameter $\sigma_{AX,ff}^2$ to account for the additive effect of the X chromosome in both sexes: $L_f = \sigma_{AX,ff}/\sigma_{AX,ff}$ ($= 1$) for females and $L_m = \sigma_{AX,mm}/\sigma_{AX,ff}$ for males.

No dosage compensation—Without dosage compensation, $\mu_{AX,f} = 2\mu_{AX,m}$ and $\sigma_{AX,ff}^2 = 2\sigma_{AX,mm}^2$; consequently, $L_m = 1/2$. Then,

$$\begin{aligned} \text{cov}_X(\text{female, female}) &= \Psi_{ff}(2)(1)^2 \sigma_{AX,ff}^2 = 2\Psi_{ff} \sigma_{AX,ff}^2, \\ \text{cov}_X(\text{male, male}) &= \Psi_{mm}(1/\sqrt{2})^2 \sigma_{AX,ff}^2 = (1/2)\Psi_{mm} \sigma_{AX,ff}^2, \\ \text{cov}_X(\text{male, female}) &= \Psi_{mf}(\sqrt{2})(1/\sqrt{2})(1) \sigma_{AX,ff}^2 = \Psi_{mf} \sigma_{AX,ff}^2 \\ &\quad (\text{taking } \rho_{Xg,mf} = 1). \end{aligned}$$

Inclusion of sex as a covariate is necessary for analyses using this model, since we expect different means in males and females [Bulmer 1985, Ekström 2004].

Dosage compensation—With dosage compensation, the mean additive genetic effect of X-linked loci is the same in both sexes, but $\sigma_{AX,ff}^2$ is half of $\sigma_{AX,mm}^2$; $L_m = 2$. Then,

$$\begin{aligned}\text{cov}_x(\text{female, female}) &= \Psi_{ff}(2)(1)^2\sigma_{AX,ff}^2 = 2\Psi_{ff}\sigma_{AX,ff}^2, \\ \text{cov}_x(\text{male, male}) &= \Psi_{mm}(\sqrt{2})^2\sigma_{AX,ff}^2 = 2\Psi_{mm}\sigma_{AX,ff}^2, \\ \text{cov}_x(\text{male, female}) &= \Psi_{mf}(\sqrt{2})(1)\sigma_{AX,ff}^2 = 2\Psi_{mf}\sigma_{AX,ff}^2.\end{aligned}$$

All coefficients in $\mathbf{X}$ are equal to 2, so Equation 4 simplifies to

$$\text{cov}_x(i, j) = 2\Psi\sigma_{AX,ff}^2. \quad (5)$$

These results are summarized in Table 2.

SIMULATION EXPERIMENTS

POWER AND ACCURACY

All simulations and analyses were carried out using the genetic analysis software SOLAR [Almasy & Blangero 1998]. For initial calculations of power and accuracy, we simulated a normally distributed trait whose sole source of additive genetic variance was an X-linked QTL with two alleles at equal frequencies. Hardy-Weinberg equilibrium was assumed. The underlying pedigree structure, based on a subset of the San Antonio Family Heart Study [Mitchell et al. 1996], is representative of many large family studies. It has a slightly female-biased sex ratio (658 females: 622 males).

We modified the autosomal QTL simulation routine within SOLAR as follows: Female phenotypic means associated with the QTL genotypes were calculated as for an autosomal QTL across a range of heritabilities (0–70%). Males were required to inherit two copies of the same QTL allele. For a trait modeled assuming dosage compensation, the phenotypic mean associated with each male “homozygote” was set equal to the corresponding female homozygote. For a trait modeled assuming the absence of dosage compensation, the male genotype producing the lower value of the trait was given a phenotypic value equal to that of the corresponding female homozygote, while the other was set equal to the female heterozygote (Table 3). The male/female ratios of phenotypic means and variances in these simulations approached the predictions of Bulmer [1985] as the proportion of total variance due to the QTL approached unity (Figure 1).

We ran a series of simulations (1000 replicates per experiment) to compare the power and accuracy of estimation of the effect of an X-linked QTL simulated either with or without dosage compensation. We analyzed the simulated data in two ways: assuming dosage compensation (DC; Equation 5) or assuming no dosage compensation (NDC; Equation 4, using the L coefficients from Table 2). In each replicate, the null hypothesis was a polygenic model (all additive genetic effect attributed to autosomes), estimated with separate environmental variance terms for males and females; these components were retained in the X-effect model. (Consequently, in Fig. 1 and subsequent usage, “heritability” refers to ratios of *female* variance components, since our covariance models use $\sigma_{AX,ff}^2$ as the point of reference. Thus, $h_x^2 = \sigma_{AX,ff}^2 / (\sigma_{AX,ff}^2 + \sigma_A^2 + \sigma_{Ef}^2)$, where σ_{Ef}^2 is the female environmental variance and σ_A^2 is the non-sex-specific autosomal additive genetic component. Similarly, h_r^2 is the residual (autosomal) additive genetic variance, $\sigma_A^2 / (\sigma_{AX,ff}^2 + \sigma_A^2 + \sigma_{Ef}^2)$). Inclusion of sex-specific random environmental variances is desirable to accommodate possible autosomal-genotype $\times$ sex interactions, and may increase the sensitivity to detect a significant X-effect (Kent et al., in press).

Significance was tested as twice the difference in log-likelihoods of the X-effect and null model ($2\ln\Delta$); this statistic is distributed as a 1/2:1/2 mixture of $\chi^2(df = 1)$ and a point mass at zero [Self & Liang 1987].

Power to detect an X effect in each scenario is summarized in Figure 2. All scenarios had very low false-positive rates ($\sim 0.3\%$) when the true X effect was zero. This low Type I error rate probably reflects the fact that h^2_r was simulated = 0 in these experiments; in the simulations described in the next section, where h^2_r was simulated = 0.2, the Type I error rate for the test of $h^2_x = 0$ was about 4.4%. The test that assumed dosage compensation had the greatest power to detect the effect of a trait simulated with dosage compensation (model DC); however, this same method performed poorly when the trait was simulated without dosage compensation. Except for the latter case, all methods had $>90\%$ power to detect an X effect when the heritability was 50% or greater.

Good estimates of the heritability due to the X-linked QTL were obtained when the method of analysis matched the dosage compensation state of the simulated trait (Table 4). In these cases, there was also minimal false attribution of variance to autosomal loci; that is, h^2_r was close to zero. However, misspecification of the model resulted in substantial overestimation of h^2_r .

A combined approach—Given the apparent utility of matching the analytical model to the dosage compensation state of the loci in question (but also acknowledging uncertainty about the extent of lyonization across the X chromosome), one possible approach is to apply DC first (since most loci are subject to lyonization), and then apply NDC if DC does not reject the null. While appealing, this approach runs the risk of increasing Type I error. We performed replicate simulations ($N =$ at least 1000 replicates/experiment), as follows: A dataset was simulated over a range (0 – 0.5) of heritabilities due to an X-linked QTL, with a residual [autosomal] additive genetic heritability = 0.20. If DC did not reject the null, NDC was applied to the same data, and rejection of the null was recorded if rejected by either model. In preliminary experiments, this combined method had an elevated Type I error rate ($\sim 7\%$ rejection of null) when the true X effect was zero. Consequently, the rejection threshold was lowered to a Bonferroni-corrected $p < 0.025$. At each replicate, the same data were then tested with a 3-parameter model (3P) that estimated the male and female X-linked variances and their correlation $\rho_{Xg,mf}$; the latter parameter was constrained to the interval $\{-1, 1\}$. Lyonization coefficients were fixed = 1 (Table 2).

The covariance component for male-female relative pairs was estimated explicitly as $\sigma_{AX,ff}\sigma_{AX,mm}\rho_{Xg,mf}$. This re-parameterization of the covariance term of Ekström [2004] is consistent with our derivation of the lyonization coefficients above and in the Appendix (and, for technical reasons, is more conveniently estimated by SOLAR). However, as noted by Amos et al. [2001] in the context of bivariate polygenic analysis, the appropriate asymptotic distribution of the test statistic for this case is somewhat uncertain: the correlation term is undefined (and the corresponding covariance term of [Ekström 2004] is necessarily zero) if either variance term is 0, leaving the precise number of degrees of freedom in doubt. We used an empirical distribution of the test statistic when the trait was simulated under the null hypothesis (i.e., no X-linked additive genetic contribution) to establish the $P_{0.05}$ critical value of the test statistic for this comparison (see Appendix).

We recorded the rate of rejection of the null hypothesis and analysis time for the two methods (DC+NDC vs. 3P). When the trait was simulated with dosage compensation, the combined DC+NDC model had comparable power even at the Bonferroni-corrected significance threshold of $p < 0.025$ (Table 5). When the trait was simulated without dosage compensation, the two models had comparable power at higher values of h^2_x ; at lower

values of h^2_x , the 3P model was somewhat more powerful. It should be noted, however, that the relative power of the 3P model is sensitive to the use of the empirically-derived $P_{0.05}$ critical value. As discussed in the Appendix, the distribution of the test statistic derived in [Ekström 2004] is probably too conservative; if that theoretical distribution is used, the 3P model has substantially lower power than DC+NDC (data not shown). In addition, there is evidently a cost in computation time for estimating additional parameters. DC+NDC had a shorter analysis time than 3P in all cases, even when both DC and NDC were applied in the combined model, and this difference increased with increasing h^2_x . Interestingly, as the magnitude of h^2_x increased, a significant X effect due to the NDC trait was identified increasingly often by the DC test, with a consequent reduction in analysis time.

EFFECT ON DETECTION OF AN AUTOSOMAL QTL

In principle, if a phenotype is influenced by both autosomal and X-linked loci, accurate specification of the effect of the X chromosome should improve the ability to detect autosomal linkages by improving the variance-components model. A negative alternative result could be that estimation of an additional parameter could reduce the efficiency of detection. To test these alternatives, we simulated a trait ($N = 1000$ replicates/experiment) with residual [autosomal] heritability $h^2_r = 0.2$, heritability due to an autosomal QTL (h^2_q) over a range of 0 – 0.3, and residual X-linked heritability (h^2_x) over a range of 0.1 – 0.3. The residual additive genetic deviations due to the X were simulated assuming dosage compensation in the same way as the residual autosomal effect (except that the deviations were computed as the Cholesky matrix of 2Ψ rather than 2Φ). The autosomal QTL was simulated with two alleles at equal frequencies, and a perfectly linked marker ($\theta = 0$) was simulated with three alleles at frequencies of 0.2, 0.3, and 0.5. In each replicate, the logarithm-of-odds score (LOD) for the QTL was estimated with and without estimation (assuming dosage compensation) of h^2_x . Estimation of the X effect did not reduce the evidence for the autosomal locus, and at higher values of h^2_q slightly increased the mean LOD score (Table 6).

CONCLUSIONS

We propose an efficient method for estimating the additive genetic contribution of the X chromosome from pedigree and phenotype data alone, even when microsatellite or other marker data are not available for this chromosome. We have expanded on the algorithm of [Ekström 2004] by explicitly modeling the effect of either presence or absence of eutherian-type dosage inactivation. Because both possibilities are tested, our method should be applicable to both lyonizing and nonlyonizing X-linked loci, and can be applied to organisms with alternate forms of dosage compensation. (There is one important exception: any case in which dosage compensation is not random with respect to the parental origins of the sex chromosomes. In marsupials, the paternal copy of the X chromosome is preferentially inactivated; this is a special case of imprinting and beyond the scope of this paper.)

Explicit modeling of the effect of dosage compensation allows us to employ Bulmer's [1985] observation of a linear relationship between the variances due to the X in males and females. Consequently, we can estimate the X effect as a single parameter, rather than for each sex separately. This has an intuitive appeal as it provides a simple measure of the relative importance of X-linked loci to a trait of interest. It also offers improvements in power when dosage compensation is present – the most common case in humans and other eutherians – as well as improvement in computation time in all cases. Compared with a 3-parameter model comparable to that of [Ekström 2004], computation time in our simulations was reduced, on average, by 33% for non-dosage compensated traits to 63% for dosage-compensated traits (Table 5). This improvement in processing time is negligible for a single

analysis but should be quite substantial when multiple analyses are required, as in multipoint whole-genome scans requiring model estimation at hundreds of loci.

Our approach offers an efficient method for estimating the residual X-linked additive genetic variance in linkage analyses, but can also test for evidence of X-linkage even when marker data is unavailable. The magnitude of the X effect will depend on the number of QTLs influencing a trait, and what subset of these resides on the X. Thus, if at least one QTL is present on this chromosome, the power to detect an additive genetic effect of the X should be no different than the power to detect an autosomal QTL – and indeed, the sensitivity of our analysis resembles that of conventional linkage analysis (MC Mahaney, pers. commun.)

Our work incorporates estimation of separate environmental variances for males and females, to allow for autosomal-genotype $\times$ sex interaction [Kent et al. in press] – or, potentially, for differences in environment due to gender roles or other cultural factors. If it is deemed preferable to estimate the male-male, female-female, and male-female variance components separately, our work suggests some adjustments to the model of [Ekstrøm 2004]. The coefficient due to allele-sharing should be $\frac{1}{2}$ rather than $\frac{1}{4}$, and the distribution of the test statistic presented by Ekstrøm may be overly conservative.

Our proposed method has been implemented within the genetic analysis software package SOLAR [Almasy & Blangero 1998].

Acknowledgments

This work was supported in part by NIH/NIMH grant MH059490-06 (J. Blangero, principal investigator). We appreciate the thoughtful comments of three anonymous reviewers. Harald H. H. Göring provided helpful advice on the computation of the empirical test statistic distribution for the three-parameter model.

References

- Abecasis, GR. [accessed 1/13/2005] MINX: Chromosome X Analyses. The University of Michigan Center for Statistical Genetics MERLIN Reference Sheet. No date. [<http://www.sph.umich.edu/csg/abecasis/Merlin/reference.html>]
- Almasy L, Blangero J. Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998; 62:1198–1211. [PubMed: 9545414]
- Amos CI, de Andrade M, Zhu DK. Comparison of multivariate tests for genetic linkage. *Hum Hered.* 2001; 51:133–144. [PubMed: 11173964]
- Amrein H. Multiple RNA-protein interactions in *Drosophila* dosage compensation. *Genome Biol.* 2000; 1:reviews1030.1–reviews1030.5. [PubMed: 11178270]
- Bulmer, MG. *The Mathematical Theory of Quantitative Genetics*. New York: Oxford University Press; 1985.
- Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005; 434:400–404. [PubMed: 15772666]
- Cordell HJ, Kawaguchi Y, Todd JA, Farrall M. An extension of the Maximum Lod Score method to X-linked loci. *Ann Hum Genet.* 1995; 59:435–449. [PubMed: 8579335]
- Ekstrøm CT. Multipoint linkage analysis of quantitative traits on sex-chromosomes. *Genet Epidemiol.* 2004; 26:218–230. [PubMed: 15022208]
- Kent JW Jr, Lease LR, Mahaney MC, Dyer TD, Almasy L, Blangero J. X chromosome effects and their interactions with mitochondrial effects. In press.
- Grossman M, Eisen EJ. Inbreeding, coancestry, and covariance between relatives for X-chromosomal loci. *J Hered.* 1989; 80:137–142. [PubMed: 2926116]
- Lyon MF. Gene action in the X chromosome of the mouse (*Mus musculus* L.). *Nature.* 1961; 190:372–373. [PubMed: 13764598]

- Lynch, M.; Walsh, B. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc; 1998.
- Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp M, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans: the San Antonio Family Heart Study. *Circulation*. 1996; 94:2159–2170. [PubMed: 8901667]
- Pérez-Enciso M, Clop A, Folch JM, Sánchez A, Oliver MA, Óvilo C, Barragán C, Varona L, Noguerra JL. Exploring alternative models for sex-linked quantitative trait loci in outbred populations: Application to an Iberian × Landrace pig intercross. *Genetics*. 2002; 161:1625–1632. [PubMed: 12196405]
- Self SG, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc*. 1987; 82:605–610.
- Stram DO, Lee JW. Variance components testing in the longitudinal effects mixed model. *Biometrics*. 1994; 50:1171–1177. [PubMed: 7786999]

APPENDIX

1. HETEROGENEOUS LYONIZATION

In the case where λ_q is constant, the relationship between male and female variances for the phenotype of q is as developed above (Table 1). However, recent evidence [Carrell & Willard 2005] suggests that lyonization may be variable among human females at ~10% of X-linked loci. In this most general case, two females with the same genotype at q may express the genotype differently, influencing the correlation between their phenotypic values. For the sake of completeness, a generalized model allowing for heterogeneous lyonization is sketched below; appropriate simulation tests of the proposed model will be presented in a future report.

Let $\{(i,j)_{k1}, (i,j)_{k2}, \dots, (i,j)_{kn}\}$ be the set of all pairs of females in a population who share the same genotype k at q . Let $(1 + \lambda_{q,i})/2$ be the mean expressivity due to lyonization at q for individual i . Then the covariance between expressed values of the genotypes in pairs of individuals can be calculated as,

$$\text{cov}_{X_{qf},ff} = \sum_k f_k \sum_{i,j} ((1 + \lambda_{q,i})(0.5)q_k - Q)((1 + \lambda_{q,j})(0.5)q_k - Q),$$

where f_k = the frequency of genotype k at q , q_k = the additive genetic contribution of that genotype in the absence of lyonization, and Q is the mean expressed value (across all genotypes) at q . Recognizing that individuals i, j are drawn with replacement from a single population of females, we can write the correlation of expressed values as,

$$\rho_{X_{qf},ff} = \left[\sum_k f_k \sum_{i,j} ((1 + \lambda_{q,i})(0.5)q_k - Q)((1 + \lambda_{q,j})(0.5)q_k - Q) \right] / \sigma_{AX,ff}^2.$$

In practice, locus-specific information (genotype means, allele frequencies, etc.) is usually unavailable. However, the locus-specific correlations may be averaged across all loci on the X-chromosome (the average correlation, of course, being influenced by the number of X-linked QTLs affecting the trait and the lyonization state at each QTL). The average correlation may be estimated as part of the covariance decomposition for female-female pairs:

$$\text{cov}_X(\text{female, female}) = \Psi_{ff}(2) \rho_{Xg,ff} \sigma_{AX,ff}^2,$$

where the lyonization coefficients are replaced by the correlation term. Note that when λ_q is constant, $\rho_{Xg,ff} = 1$.

For male-female pairs,

$$\rho_{Xgq,mf} = \left[\sum_k f_k \sum_{i,j} ((m_{q,k} q_k - Q)((1 + \lambda_{q,j})(0.5)q_k - Q)) \right] / \sigma_{AX,ff}^2.$$

This is a weighted correlation, in that the denominator contains $\sigma_{AX,ff}^2$ rather than $\sigma_{AX,mm} \sigma_{AX,ff}$; this is required since the X effect is expressed in terms of $\sigma_{AX,ff}^2$. It also contains an X-genotype $\times$ sex interaction term $m_{q,k}$ for the male partner. Unlike $\lambda_{q,i}$, this term can take any nonzero value (null expectation = 1), and it is constant for all males of like genotype. The expression for the correlation becomes singular when $m_{q,k} \rightarrow 0$ or $m_{q,k} \rightarrow \infty$ (or equivalently, when one sex-specific variance is zero while the other has a nonzero value). Thus, the model can handle moderate X-genotype $\times$ sex interaction but should not be applied to a case in which an X-linked gene is expressed only in one sex.

The resulting decomposition of the male-female covariance (again averaging across loci) is,

$$\text{cov}_X(\text{male, female}) = \Psi_{ff}(\sqrt{2}) \rho_{Xg,mf} \sigma_{AX,ff}^2,$$

which expands the interpretation of the male-female correlation term introduced in Equation 3b to include both X-genotype $\times$ sex interactions and dosage compensation.

The null models nested within this general model are:

- a. $\rho_{Xg,mf} = (\sigma_{AX,mm} / \sigma_{AX,ff}) \rho_{Xg,ff}$ ($m_{q,k} = 1$ in the absence of X-genotype $\times$ sex interaction), and
- b. $\rho_{Xg,ff} = 1$ (in the absence of *heterogeneous* lyonization).

Each null model has one less parameter than its more general model, which can be tested for significance with a 1 d.f. likelihood-ratio test. $m_{q,k}$ is tested within its range, while $\rho_{Xg,ff}$ is tested on its upper boundary. Thus the test statistic (twice the difference in loglikelihoods) for $m_{q,k}$ is distributed as χ^2 with d.f. = 1, while the test statistic for $\rho_{Xg,ff}$ is distributed as a 1/2:1/2 mixture of χ^2_0 and χ^2_1 [Self & Liang 1987].

It follows that the male-male covariance is decomposed as,

$$\text{cov}_X(\text{male, male}) = \Psi_{ff}(\sqrt{2}) (\rho_{Xg,mf})^2 \sigma_{AX,ff}^2.$$

2. EMPIRICAL DISTRIBUTION OF TEST STATISTIC FOR MODEL 3P

In the case where the male and female X-linked variances and correlation are estimated separately, the experimental and null hypotheses are:

$$\begin{aligned}
 H_1: \Omega &= 2\Phi\sigma_A^2 + \mathbf{I}\sigma_E^2 + 2\Psi_{ff}\sigma_{AX,ff}^2 + \Psi_{mm}\sigma_{AX,mm}^2 + \sqrt{2}\Psi_{mf}\sigma_{AX,ff}\sigma_{AX,mm}\sigma_{Xg,mf} \\
 H_0: \Omega &= 2\Phi\sigma_A^2 + \mathbf{I}\sigma_E^2
 \end{aligned}$$

H_1 is a re-parameterization of the 3-parameter model of Ekstrøm [2004] (which uses the covariance rather than the correlation). In developing the distribution of the test statistic $2\ln\Delta$ under H_0 , Ekstrøm notes that the variance terms are constrained to their respective lower boundaries ($= 0$), while the covariance, constrained to 0, is in the middle of its range. He therefore computes a mixture of chi-squares with 1, 2, and 3 degrees of freedom by the method of [Self & Liang 1987]. However, as noted by Amos et al. [2001] for a similar case (a bivariate polygenic model), the appropriate distribution for this case may be hard to define: the correlation in our model 3P is undefined if either variance term is 0, and the covariance in Ekstrøm's model is necessarily 0 (and thus not independent) in these cases.

We have chosen to simulate a trait under the null hypothesis of no X-linked additive genetic contribution to obtain an empirical distribution of the test statistic for 3P. A trait with 20% of its variance due to autosomal additive genetic effects was simulated in 21,168 replicates using the same underlying pedigree structure as the other simulations in this study (an additional 12,000 replicates with $h^2_r = 0.4$ yielded similar results, not shown). $2\ln\Delta$ was computed versus H_0 for both the 1-parameter dosage compensation model (DC) and the 3-parameter model in each replicate. The 90th, 95th, and 99th percentile values of these empirical distributions, corresponding to $\alpha = 0.10, 0.05$, and 0.01 , are shown in Table A.1. The values for the 1-parameter model agree closely with those from the expected mixed distribution ($1/2:1/2 \chi^2_0: \chi^2_1$) [Self & Liang 1987]. The empirical critical values for 3P, however, are roughly equivalent to those expected from a mixed distribution for *two* new parameters tested on their boundaries: $1/4:1/2:1/4 \chi^2_0: \chi^2_1: \chi^2_2$.

In the above comparison of the 1- and 3-parameter models, we used the empirical critical value $P_{0.05} = 4.3$ to test the significance of 3P. An additional consequence – since our 3P model and Ekstrøm's model yield the same likelihoods given the same data – is that the null distribution of [Ekstrøm 2004] may be overly conservative.

Figure 1. Male/female ratios of means and variances for a simulated trait

Panel A, male/female ratio of phenotypic means, and panel B, ratio of phenotypic variances, for a trait simulated across a range (0.1 – 0.98) of heritabilities due to an X-linked QTL and $h^2_r = 0$. The trait was simulated either with dosage compensation for the X (closed circles) or without dosage compensation (open squares).

Figure 2. Power to detect an effect of the X chromosome

Power to detect an X effect estimated as percent rejection of the null hypothesis (no additive genetic effect of the X) for a trait simulated over a range (0 – 0.7) of heritabilities due to an X-linked QTL and $h^2_r = 0$. $N = 1000$ replicates/experiment. Trait was simulated with dosage compensation (closed circles) or without dosage compensation (open squares). X effect was tested assuming dosage compensation (solid lines) or assuming no dosage compensation (broken lines).

Table 1
Phenotypes, means, and variances of a trait influenced by a diallelic X-linked locus with constant inactivation

Male genotype	Phenotype	Frequency
A1	1	p
A2	0	1-p
Mean = p(1) + (1-p)(0) = p		
Variance = p(1-p) ² + (1-p)(0-p) ² = p(1-p)		
General case:		
Female genotype	Phenotype	Frequency
A1A1	[(1 + λ)/2](2) = 1 + λ	p ²
A1A2	[(1 + λ)/2](1) = (1 + λ)/2	2p(1-p)
A2A2	[(1 + λ)/2](0) = 0	(1-p) ²
Mean = p ² (1 + λ) + 2p(1-p)(1 + λ)/2 + (1-p) ² (0) = (1 + λ)p		
Variance = p ² [(1 + λ) - (1 + λ)p] ² + 2p(1-p)[(1 + λ)/2 - (1 + λ)p] ² + (1-p) ² [0 - (1 + λ)p] ²		
= p(1-p)(1 + λ) ² /2		
Limit case: no dosage compensation (λ = 1 at all loci)		
Female mean = 2p (i.e., twice the male mean)		
Female variance = 2p(1-p) (i.e., twice the male variance)		
Limit case: with dosage compensation (λ = 0 at all loci)		
Female mean = p (i.e., equal to the male mean)		
Female variance = p(1-p)/2 (i.e., half the male variance)		

See text for explanation. The subscript of λ_q is omitted for clarity.

Table 2

Structuring elements and variance components for X-linked additive genetic covariances

General case:						Dosage compensation:		No dosage compensation:	
Ψ_{ij} :	I_{ij} :	L_{ij} :	Variance:	L_{ij} :	Variance:	L_{ij} :	Variance:	L_{ij} :	Variance:
Female-female	$P_1/4 + P_2/2$	2	$\sigma^2_{AX,ff}$	1	$\sigma^2_{AX,ff}$	1	$\sigma^2_{AX,ff}$	1	$\sigma^2_{AX,ff}$
Male-male	P_1	1	$\sigma^2_{AX,mm}$	2	$\sigma^2_{AX,ff}$	1/2	$\sigma^2_{AX,ff}$		
Male-female	$P_1/2$	$2\rho_{Xg,mf}^*$	$\sigma_{AX,mm}\sigma_{AX,ff}$	2	$\sigma^2_{AX,ff}$	1/2	$\sigma^2_{AX,ff}$		

* In the absence of X-genotype by sex interactions (or variable lyonization, see Appendix), $\rho_{Xg,mf} = 1$.

Table 3

Genotype-specific phenotypic means for simulated traits

With dosage compensation		Without dosage compensation	
Females;			
genotype:	<i>aa</i>	<i>bb</i>	<i>bb</i>
phenotype:	100- <i>d</i>	100	100+ <i>d</i>
Males;			
genotype:	<i>a</i>	<i>b</i>	<i>b</i>
phenotype:	100- <i>d</i>	100+ <i>d</i>	100

Mean phenotypic effect (arbitrary units) of specified genotypes for an X-linked QTL simulated with or without dosage compensation. The deviation *d* was calculated for each given QTL heritability assuming a normally distributed trait, within-genotype standard deviation = 10 units, and Hardy-Weinberg equilibrium.

Table 4

Heritability estimates for simulated traits

Simulated with dosage compensation				
Simulated h^2_x:	DC estimates		NDC estimates	
	h^2_x	h^2_r	h^2_x	h^2_r
0.00	0.00	0.01	0.00	0.01
0.10	0.08	0.03	0.10	0.05
0.20	0.18	0.03	0.23	0.08
0.30	0.28	0.03	0.33	0.11
0.50	0.48	0.03	0.51	0.18
0.70	0.68	0.04	0.65	0.25

Simulated without dosage compensation				
Simulated h^2_x:	DC estimates		NDC estimates	
	h^2_x	h^2_r	h^2_x	h^2_r
0.00	0.00	0.00	0.01	0.01
0.10	0.03	0.04	0.08	0.02
0.20	0.05	0.08	0.19	0.02
0.30	0.08	0.13	0.29	0.02
0.50	0.11	0.24	0.50	0.02
0.70	0.14	0.38	0.70	0.02

Point estimates are given for the h^2_x and [spurious] h^2_r for an X-linked trait simulated with and without dosage compensation (standard errors for all estimates are < 0.005). Heritabilities were estimated from models that used lyonization coefficients appropriate for dosage compensation (DC) or no dosage compensation (NDC). $N = 1000$ replicates/experiment.

Table 5

Comparison of combined 1-parameter test vs. 3-parameter test

Test: DC+NDC					Test: 3P	
Simulated h^2_x	% rejection of null	% times rej. by DC	runtime, s mean (SD)	% rejection of null	runtime, s mean (SD)	
0.00	4.6	54.4	32.4 (8.4)	11.3	37.3 (10.1)	
<i>DC trait</i>						
0.10	60.8	94.7	23.3 (12.7)	62.2	39.3 (8.2)	
0.20	96.5	98.9	14.3 (7.2)	93.5	43.1 (9.4)	
0.30	100.0	100.0	11.8 (3.2)	98.5	40.4 (7.4)	
0.50	100.0	100.0	11.2 (3.0)	99.7	43.4 (9.7)	
<i>NDC trait</i>						
0.10	31.8	32.1	36.5 (10.7)	57.2	42.4 (10.4)	
0.20	81.6	32.3	30.3 (11.7)	92.8	39.1 (8.8)	
0.30	98.3	42.9	25.9 (13.4)	100.0	39.9 (7.3)	
0.50	100.0	83.6	14.3 (9.6)	100.0	38.1 (5.9)	

Power to detect an X effect estimated as percent rejection of the null hypothesis (no additive genetic effect of the X) for a trait simulated over a range (0 – 0.5) of heritabilities due to an X-linked QTL and $h^2_T = 0.2$. $N =$ at least 1000 replicates/experiment. Trait was simulated with dosage compensation (DC) or without dosage compensation (NDC). For each replicate, X effect was estimated by a combined 1-parameter test (DC+NDC) and subsequently by a 3-parameter test (3P).

Table 6

Effect of estimating X effect on detection of an autosomal QTL

Simulated			Observed			
h^2_q	h^2_x	LOD mean (SE), h^2_x not est.	LOD mean (SE), h^2_x est.	% rej. of null, h^2_x not est.	% rej. of null, h^2_x est.	
0.0	0.3	0.112 (0.007)	0.108 (0.007)	5.5	4.8	
0.1	0.1	0.798 (0.024)	0.800 (0.024)	49.6	50.2	
0.1	0.2	0.886 (0.027)	0.899 (0.027)	53.4	54.7	
0.1	0.3	0.970 (0.029)	0.999 (0.030)	57.1	58.8	
0.2	0.1	2.539 (0.050)	2.541 (0.050)	92.8	92.6	
0.2	0.2	2.745 (0.051)	2.782 (0.051)	94.1	94.4	
0.2	0.3	3.007 (0.054)	3.094 (0.054)	95.8	96.0	
0.3	0.1	5.334 (0.071)	5.338 (0.071)	99.8	99.8	
0.3	0.2	5.713 (0.077)	5.774 (0.077)	99.8	99.8	
0.3	0.3	6.219 (0.081)	6.334 (0.082)	99.9	99.8	

$N = 1000$ replicates/test; $h^2_T = 0.2$ in all tests.

Table A.1Empirical and theoretical critical values for $2\ln\Delta$

Percentile	1-X-linked parameter model		3-X-linked parameter model	
	Empirical	Theoretical ¹	Empirical	Theoretical ²
0.90	1.624	1.642	2.952	2.952
0.95	2.670	2.706	4.298	4.230
0.99	5.430	5.412	7.452	7.288

¹ $_{1/2:1/2} \chi^2_0: \chi^2_1$ ² $_{1/4:1/2:1/4} \chi^2_0: \chi^2_1: \chi^2_2$