Published in final edited form as:

Genet Epidemiol. 2013 April ; 37(3): 239–247. doi:10.1002/gepi.21706.

Association Testing of the Mitochondrial Genome Using Pedigree Data

Chunyu Liu^{1,2,*}, Josée Dupuis³, Martin G Larson⁴, and Daniel Levy^{1,2}

¹National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts 01702, USA

²Center for Population Studies, NHLBI, Bethesda, Maryland 20892, USA

³Department of Biostatistics, School of Public Health, Boston University, Massachusetts 02118, USA

⁴Department of Mathematics and Statistics, Boston University, Massachusetts 02215, USA

Abstract

In humans, mitochondria contain their own DNA (mtDNA) that is inherited exclusively from the mother. The mitochondrial genome encodes thirteen polypeptides that are components of oxidative phosphorylation to produce energy. Any disruption in these genes might interfere with energy production and thus contribute to metabolic derangement. Mitochondria also regulate several important cellular activities including cell death and calcium homeostasis. Aided by sharply declining costs of high-density genotyping, hundreds of mitochondrial variants will soon be available in several cohorts with pedigree structures. Association testing of mitochondrial variants with disease traits using pedigree data raises unique challenges because of the difficulty in separating the effects of nuclear and mitochondrial genomes, which display different modes of inheritance. Failing to correctly account for these effects might decrease power or inflate type I error in association tests. In this report, we sought to identify the best strategy for association testing of mitochondrial variants when genotype and phenotype data are available in pedigrees. We proposed several strategies to account for polygenic effects of the nuclear and mitochondrial genomes and we performed extensive simulation studies to evaluate type I error and power of these strategies. In addition, we proposed two permutation tests to obtain empirical p-values for these strategies. Furthermore, we applied two of the analytical strategies to association analysis of 196 mitochondrial variants with blood pressure and fasting blood glucose in the pedigree rich, Framingham Heart Study. Finally, we discussed strategies for study design, genotyping, and data cleaning in association testing of mtDNA in pedigrees.

Keywords

mitochondrial DNA; association test; polygenic effect; maternal lineage; permutation test

*Corresponding author: Chunyu Liu, PhD, Framingham Heart Study, 73 Mt. Wayte Avenue, Suite 2, Framingham, MA 01702, chunyu.liu@nih.gov, Fax: 508-872-2678, Telephone: 508-663-4057.

The authors declare that there is no conflict of interest.

Introduction

Mitochondria are the energy powerhouses of the cell. In humans, the three major nutrients -- carbohydrates, proteins, and fats - are broken down into simpler molecules that are used for energy production through the process of oxidative phosphorylation (OXPHOS) [Voet, et al. 2006]. Therefore, mitochondria are a converging point of these major metabolic pathways. Mitochondria contain their own DNA (mtDNA) that exists in a circular molecule of approximate 16.6 kb. Mitochondrial DNA encodes thirteen polypeptides that are essential components of OXPHOS. Any disruption in these genes might adversely affect energy production and contribute to metabolic derangements. The metabolic syndrome is defined by the coexistence of multiple metabolic risk factors, such as obesity, a high triglyceride level, high blood pressure (BP), high fasting blood glucose (FBG), and low high density lipid (HDL). These factors raise the risk for heart disease, diabetes, and stroke. In addition to their role in energy production, mitochondria regulate cell death pathways [Green 1998], calcium homeostasis [Hajnoczky, et al. 2006], cellular proliferation and intermediate metabolism [McBride, et al. 2006], and biosynthesis of steroid [Oh-hama 1997] and certain heme molecules [Rossier 2006]. Mitochondrial DNA also encodes 12S and 16S rRNA and 22 tRNA genes that are required for mitochondrial protein synthesis.

Mitochondrial DNA is maternally inherited and lacks recombination [Elson, et al. 2001]. This feature enables researchers to trace maternal line far back in time based on variations in the mitochondrial genome. Some unique combination of common mtSNPs can be used to define large haplogroups [Torroni, et al. 1994]. Among people of European ancestry, 95% belong to one of nine haplogroups [Torroni, et al. 1996]. Several studies have investigated the association between mtDNA haplogroups and complex diseases [Benn, et al. 2008; Herrnstadt and Howell 2004; Saxena, et al. 2006; van der Walt, et al. 2004]. Single common mtSNPs have also been used in association analyses of several common diseases [Byrne, et al. 2009; Kim, et al. 2002; Lal, et al. 2005; Okura, et al. 2003; Poulton, et al. 2002; Saxena, et al. 2006]. Findings from these studies have rarely been replicated in either haplogroup or single mtSNP association analyses, perhaps due to lack of power [Samuels, et al. 2006] or due to the fact that only common mtSNPs have been investigated in most of previous studies.

Heteroplasmy refers to a state where more than one mitochondrial genome exists within a cell, tissue, or individual, and it can be used as genetic marker in several applications [He, et al. 2010]. Heteroplasmy can be inherited from the mother or occur *de novo* in offspring [Goto, et al. 2011; Sondheimer, et al. 2011]. Although originally believed to be a rare phenomenon, recent re-sequencing efforts yielded inconsistent findings [Goto, et al. 2011; He, et al. 2010; Li, et al. 2010].

Rare mutations in mtDNA can lead to a number of inherited diseases [Taylor and Turnbull 2005], including Kearns-Sayre syndrome (KSS) [Holt, et al. 1988] and Leber's hereditary optic neuropathy (LHON) [Wallace, et al. 1988]. Because mitochondria are a vital organelle within all cell types, mtDNA diseases affect many tissues and display various ranges of clinical features [Taylor and Turnbull 2005]. Recently, studies have shown that mtDNA mutations may be involved in the development of common diseases, such as hypertension

[Li, et al. 2009; Liu, et al. 2009; Wang, et al.; Wang, et al. 2011; Wilson, et al. 2004] and diabetes [Liu, et al. 2012; Petersen, et al. 2003; Poulton, et al. 2002]. A number of genetic risk factors (common single nuclear polymorphisms or SNPs) have been identified for these common diseases. However, the nuclear genome susceptibility loci identified to date explain only a small proportion of the heritability for most common diseases [Maher 2008]. Thus the search for the “missing heritability” has become an important task for the human genetics community and mtDNA may be part of the equation.

Aided by sharply declining costs of high-density genotyping, hundreds of mitochondrial variants will soon be available in several cohorts with pedigree structures. Testing for association of mitochondrial variants with traits in pedigrees raises unique challenges because of the difficulty in separating the effects of the nuclear and mitochondrial genomes [Schork and Guo 1993], which display different modes of inheritance. Failing to account for these effects correctly might decrease power or give rise to elevated type I error in association tests. In this report, we sought to identify the best strategy for association testing of individual mitochondrial variants when genotype and phenotype data are available. We explored several strategies to account for polygenic effects of the nuclear and mitochondrial genomes and we performed extensive simulation studies to evaluate type I error and power of these strategies. We determined that the best strategy is a mixed effect model that accounts for random effects in short-span maternal lineages (defined below). In addition, we propose a permutation test that can be used to obtain empirical p-values for several strategies. Our results provide guidance for association testing of mitochondrial variants in pedigrees.

Material and Methods

Defining Short-span Maternal Lineage in Pedigrees

For any multigenerational cohort, pedigree structure can be constructed and used to define short-span maternal lineages (SSMLs) described as follows. Founders and married-in individuals are assigned distinct numbers to denote their SSML. Female founders and married-in females share a common number with their respective offspring and the descendants of female offspring. In samples of nuclear families which include only both parents and their offspring, the mother and all her offspring belong to the same SSML, while the father belongs to a different SSML. Please note that SSML is a special case of haplogroups. All individuals in a SSML possess the identical mtDNA (in the absence of *de novo* mutations) while in haplogroups, individuals share some common combinations of mtSNPs.

Genetic Model

We assume that a quantitative trait is affected by several components [Schork and Guo 1993] [Amos 1994].

$$Y_i = \mu + \gamma g_i + M_i + N_i + \sum_{k=1}^S \beta_k Z_{ik} + e_i \quad (\text{Equation 1})$$

In equation 1, let Y_i represents the trait value for the i^{th} individual and μ be the overall trait mean. We assume that the continuous trait Y_i is influenced by three genetic components: a fixed genetic effect, denoted by γ (g_i is the observed genotype and γ is the fixed effect coefficient), a polygenic component, M_i , of the mitochondrial genome, and a polygenic component, N_i , of the nuclear genome. Because individuals in pedigrees are correlated due to either maternal or nuclear inheritance (or both), M_i and N_i are considered as random effects in Equation 1 to account for such correlations. Therefore, the genetic model in Equation 1 is a mixed effects model.

Let Z_{ik} be the k^{th} covariate for person i and β_k , be the effect from this covariate that is unrelated with the genetic factors. Let e_i be the residual variability that is uncorrelated with the genetic factors and covariates. We further assume $E(M_i) = E(N_i) = E(e_i) = 0$ [Amos 1994]. We assume that a mitochondrial SNP has two possible genotypes A and B, with effects of

$$g_i = \begin{cases} -\frac{1}{2} & \text{if individual } i \text{ has unobserved genotype A} \\ \frac{1}{2} & \text{if individual } i \text{ has unobserved genotype B} \end{cases} \quad (\text{Equation 2})$$

Let p and q be the genotype frequencies of A and B, respectively ($p + q = 1$).

Let σ_g^2 , σ_G^2 and σ_N^2 be the variance for the three genetic components g_i , M_i and N_i , respectively; and σ_e^2 be the random environmental effects. The variance of Y_i can be written as

$$\sigma_T^2 = \sigma_M^2 \Lambda + \sigma_N^2 \Phi + \sigma_e^2 \mathbf{I} \quad (\text{Equation 3})$$

where $\sigma_M^2 = \sigma_g^2 + \sigma_G^2$, Φ is the $N \times N$ kinship coefficient matrix (Figure 1A), and Λ is a matrix such that its ij^{th} component is 1 if pedigree members i and j are in the same SSML and is 0 otherwise (Figure 1B). From Equation 2, we obtain the variance of g as

$$\sigma_g^2 = E(\gamma g_i)^2 - (E(\gamma g_i))^2 = \gamma^2 pq. \quad (\text{Equation 4})$$

Analytical Strategies for Association Testing

In this study, we focused on quantitative traits. We considered five analytical strategies in association testing of mtDNA variants using family data, where polygenic effects were from both mitochondrial and nuclear genomes. In an association test, we tested the null hypothesis $H_0: \gamma = 0$ versus $H_1: \gamma \neq 0$ (Equation 1). The five strategies considered were as follows: 1) UI - unrelated individuals (i.e., founders or singletons) were analyzed using simple linear regression. 2) NF - all individuals were included and familial correlation (Figure 1A) due to autosomal polygenic component among nuclear families was accounted for in a linear mixed effects (LME) model. 3) MO - fathers were excluded from the sample, but mother and offspring were used for analysis. This reduced covariance structure (without

father in Figure 1A or 1B) could be accommodated by LME. 4) ML - all individuals were used. Random effect in SSML (Figure 1B) was accounted for in a LME model. 5) BO - all individuals were used. Both coefficient structures of nDNA (Figure 1A) and mtDNA (Figure 1B) were accounted for in a LME model.

Statistical Analysis

We used R version 2.9.2 (<http://cran.r-project.org/>) for all statistical analysis. We employed simple linear regression in analyzing unrelated individuals (the UI strategy) and linear mixed effects (LME) models for NF, ML, BO and MO strategies. The R function “lme4” in “lme4” package was used for NF and BO strategies. The R function “lmer” from the “lme4” package was used for ML. For MO, “lme4” and “lmer” provided identical results. We used “lme4” in this study.

Type I Error and Power Estimations

For simplicity, we set $\sigma_T^2=1$ and let $h_K^2=\sigma_K^2/\sigma_T^2$, where K represents g , G , N or e . For each genetic model, the null distribution was generated using $h_g^2=0\%$ with 200,000 iterations, and type I error rates were estimated at $\alpha = 0.001$. We estimated the empirical threshold corresponding to $\alpha = 0.001$ for each genetic model using two permutation methods described in the following section. The empirical threshold estimated from PERM_{F-M} is referred as $\alpha_{\text{PERM}_{F-M}0.001}$. Similarly, the empirical threshold estimated from PERM_M is referred as $\alpha_{\text{PERM}_M0.001}$. Power was estimated using these asymptotic and empirical thresholds.

Permutation Test

Multiple testing is an issue in association tests of multiple markers and multiple phenotypes. Due to a smaller genomic size, mtDNA harbors many fewer genetic markers than nDNA, making permutation testing feasible. In this work, we propose two ways to perform permutation tests. For both permutation methods, we performed 200,000 iterations to evaluate type-I error and to obtain empirical thresholds for power estimation. The first permutation method takes advantage of the simple structure of nuclear families. We treat a father and a mother in a nuclear family as a pair and the genotypes/haplotypes of father-mother pairs are permuted together (F-M pair permutation or PERM_{F-M}), while family structures remain the same. After each permutation, all offspring in a nuclear family will obtain the same genotype from the mother. With this permutation method, the heritability for a trait is well preserved.

However, it is infeasible to apply the PERM_{F-M} method to data that includes extended pedigrees along with singletons. We develop a second permutation method (referred as maternal permutation or PERM_M) that utilizes the special property of SSMLs, which are considered as independent units. Individuals in the same SSML share the same genotype/haplotype (assuming no *de novo* mutations). For this reason, the genotypes/haplotypes are shuffled across all SSMLs to break the relation between Y and the mitochondrial variant. After each permutation, all individuals in the same SSMLs are assigned the same permuted genotypes/haplotypes.

Simulations

We performed a simulation study to compare the five analytical strategies in terms of the type I error and power. We assumed that a quantitative phenotype of interest (Y) was affected by a major mitochondrial locus and the presence of polygenic effects from both the mitochondrial and nuclear genomes (Equation 1). In the initial simulations, we set $h_g^2=1\%$ and simulated 400 nuclear families of two parents and three offspring. We considered the minor genotype frequency (MGF) $p = 5\%$ and 10% . To evaluate how polygenic effects from both genomes affect type I error and power for these analytical strategies, we considered the following two general settings for the unknown polygenic effects: 1) the total polygenic effects were fixed at 50% (i.e., $h_G^2 + h_N^2 = 50\%$), while h_G^2 increased from 0% to 50% (Table 1A); 2) the polygenic effect from mtDNA was fixed at 15% (i.e., $h_G^2 = 15\%$) and h_N^2 was increased from 0 to 35% (Table 1B).

In secondary simulations, we conducted power analysis to calculate the sample size required to detect an effect of a given size for a mitochondrial variant when the study sample contains pedigrees. Based on results from the initial simulations, we set $h_G^2 = 15\%$ and $h_N^2 = 35\%$, and we considered $h_g^2 = 0.5\%$ and 1% and MGF = 2.5% and 5% . We evaluated power using simulations for sample sizes varying from 2,000 to 10,000 individuals, by increments of 1,000 individuals. We considered two study settings. The first setting contained only nuclear families of two parents and three offspring (option I in Table 2). The second setting was a mixture of unrelated individuals and nuclear families of different sizes (option II in Table 2). For example, in a sample of 2,000, we included 270 unrelated individuals, 250 trios, 100 nuclear families of size five (two parents and three siblings) and 60 nuclear families of size eight (two parents and six siblings). The ratio of unrelated, trios and nuclear families of size five and eight remained constant when the total sample sizes increased (Table 2).

Real Data Application

In 1948, the Framingham Heart Study (FHS) was initiated to identify risk factors for cardiovascular disease. Over time, three cohorts have been recruited: the Original cohort [Dawber, et al. 1963], the Offspring cohort that includes children and spouses of the Original cohort participants [Kannel, et al. 1979], and the Third generation cohort that are grandchildren of the Original cohort [Splansky, et al. 2007]. All study participants gave written informed consent for genetic research. All protocols in this study were approved by the Institutional Review Board of Boston University Medical Center.

Extended pedigrees were constructed from the three generational cohorts. Based on the extended pedigree structure, SSMLs were characterized in the entire FHS sample as described in the Materials and Methods section.

Genotyping, Quality Control—One hundred and ninety six mtSNPs (Supplementary Table 3) were genotyped in 2,455 Offspring participants using the Illumina *HumanOmni5*-Quad BeadChip. Pedigree and SSML structures were obtained by merging genotyped participants with the entire FHS sample. The observed frequency of heteroplasmy was very

low (please see Result) in this sample and does not provide enough power in association analysis. Thus we set all heteroplasmy cases to missing in this study.

For each mtDNA polymorphism, we compared genotypes among members of each SSML. For a SSML with three or more genotyped members, if an individual had a different genotype from all other members of the SSML, this person's genotype was set to missing. For a SSML with only two genotyped members, if their genotypes were different, both sets of genotype data were set to missing. The call rate was calculated using all individuals, while the minor allele (or genotype) frequency was calculated using one individual per SSML. In this analysis, we used a mtSNP if it existed in at least ten SSMLs, corresponding to minor allele frequency = 0.006. In all, 107 SNPs were retained for association analysis.

Inferring Genotypes and Constructing haplotypes—The SSML information was used to infer genotype for members who did not have genotyping data but were in the same SSML where other members were genotyped. We inferred genotypes in three generations corresponding to our three cohorts, but only used Offspring and the Third generation cohort data for association analysis, because previously we published results of mtDNA association using these two generations [Liu, et al. 2012]. SSMLs were also used to construct haplotypes for individuals in the same SSMLs.

Phenotypes—Continuous phenotypes included systolic and diastolic blood pressure (SBP, DBP), and fasting blood glucose (FBG). The detailed description of these phenotypes and exclusion criteria were provided previously [Dupuis, et al. 2010; Levy, et al. 2000; Levy, et al. 2009]. Briefly, the BP and FBG phenotypes at different examinations were used in analysis. BP phenotypes included a) the baseline (examination 1) trait that was measured at examination 1 for the Offspring (1971–1975) and the Third generation (2002–2005) cohorts, b) the “contemporary” trait measured at examination 7 (1998–2001) for the Offspring participants and at examination 1 (2002–2005) for the Third generation participants, and c) the “long-term averaged” BP calculated separately for SBP and DBP using mean values across examinations 1–7 for the Offspring participants. For participants receiving antihypertensive medication, adjusted SBP and DBP were calculated by adding 10 mm Hg and 5 mm Hg to the measured SBP and DBP values, respectively [Levy, et al. 2000].

The baseline FBG phenotype used measures at examination 3 from the Offspring cohort and examination 1 from the Third generation cohort. The “contemporary” phenotype consisted of FBG at examination 7 from the Offspring and examination 1 from the Third generation cohort. The long-term FBG is calculated over examinations 3–7 of the Offspring cohort.

Association Analysis and Permutation Test—The BO and ML strategies were applied to the data for association analysis between the nine phenotypes and the 107 mtSNPs. Covariates include sex, age, age-squared, body mass index, and ‘cohort’ (a variable set to 0 for Offspring participants and set to 1 for individuals from the Third generation). To account for multiple testing, the $PERM_M$ was used to evaluate empirical p-value of the ML strategy with 100,000 iterations. To preserve correlations among mtSNPs in the same phylogeny, haplotypes were shuffled across all SSMLs. The permuted genotypes can then be obtained for each subject based on the permuted haplotypes. In each iteration, association

tests were performed for all phenotypes simultaneously. The minimum p-value (minP) was obtained from 107 association analysis with a specific trait. The shuffle of the haplotypes was performed 10,000 times and association tests were performed repeatedly between permuted mtSNPs and non-permuted phenotypes, resulting in null distributions of association p-values. To correct for multiple testing for one trait, $P_{\text{empirical}}$ was calculated by comparing an observed p-value against the distribution of 100,000 values of minP. All analyses were conducted using R version 2.9.2 (<http://cran.r-project.org/>).

Results

We compare type I error rates for the five analytical strategies under the null ($H_g = 0\%$) and report empirical type I error rates at $\alpha = 0.001$ for these strategies using two permutation methods. We next compare the power for these strategies. Finally, we report association results using real mtDNA variant data genotyped on the FHS cohorts.

Type I error

Figure 2 and Supplementary Figure 1 display type I error rates (that is, $h_g^2 = 0\%$) at $\alpha = 0.001$ for the five analytical strategies when $h_G^2 + h_N^2 = 50\%$ and $h_G^2 = 15\%$, respectively. The 95% confidence interval (CI) at $\alpha = 0.001$ was plotted as well in these figures. The UI, BO and ML strategies yield correct type I error rates for all scenarios. The MO strategy has a slightly elevated type I error rate, about 30% inflation for most scenarios, regardless of the level of h_G^2 or the value of h_N^2 at a fixed level of h_G^2 . By contrast, the type I error rate for the NF strategy increases monotonically when h_G^2 increases. When $h_G^2 \leq 5\%$, the NF strategy displays a slightly elevated type I error rate (~ 0.0012). When $h_G^2 > 5\%$, the NF strategy's type I error rate becomes unacceptably high, from ~ 0.002 to 0.005 when h_G^2 increases from 15% to 50%. When h_G^2 remains constant (15%), the type I error rate appears to be at a constant level (~ 0.002) regardless of the value of h_N^2 (Supplementary Figure 1).

Next, we compare the empirical type I error rates at $\alpha = 0.001$ for these five strategies evaluated using the two permutation methods. Figure 3 and Supplementary Figure 2 provide the empirical type I error rates for the five strategies using the PERM_M method, and for the NF and BO strategies using PERM_{F-M} method when $h_G^2 + h_N^2 = 50\%$ and $h_G^2 = 15\%$, respectively. In general, both permutation methods yield correct empirical type I error for the ML, MO, and UI strategies (results not shown from PERM_{F-M}). For the BO strategy, PERM_{F-M} yields correct empirical type I error rate for all scenarios we investigated, whereas PERM_M yields a very conservative type I error rate, which becomes even more conservative when the proportion of h_G^2 increases from 0% to 50% (Figure 3) or the level of h_N^2 increases (Supplementary Figure 2). For the NF strategy, the two permutation methods yield conservative (or very conservative) empirical type I error rates for a majority of the scenarios investigated (Figure 3 and Supplementary Figure 2). The type I error rate of the NF strategy ranges from about 0.0009 to 0.0001 for PERM_{F-M} and from $4.1\text{E-}05$ to $9.9\text{E-}06$ for PERM_M as the proportion of h_G^2 is increased from 0% to 50% (Figure 3). At a fixed level

of h_G^2 , PERM_{F-M} seems to yield a type I error rate for the NF strategy at ~ 0.0004 , which is somewhat constant regardless of the value of h_N^2 (Supplementary Figure 2). The type I error rate of the BO strategy is very similar to that of the NF strategy for PERM_M (Figure 3 and Supplementary Figure 2).

Power

Figure 4 and Supplementary Figure 3 demonstrate the power for the strategies with correct empirical rates at $\alpha_{\text{PERM}_M_{0.001}}$ and $\alpha_{\text{PERM}_{F-M}_{0.001}}$ when $h_G^2 + h_N^2 = 50\%$ and $h_G^2 = 15\%$, respectively. The power for the NF strategy is not displayed because its empirical type I error is conservative using both permutation methods. For three strategies -- UI, MO, and ML -- power at $\alpha_{\text{PERM}_M_{0.001}}$ is similar to that at α_{PERM_M} ; therefore, we plot power at $\alpha_{\text{PERM}_M_{0.001}}$ in the figures. For the BO strategy, we plot power at $\alpha_{\text{PERM}_{F-M}_{0.001}}$. Among the plotted strategies, the UI strategy utilizes the unrelated individuals in the pedigree data and appears to have a constant power at a fixed level of h_g^2 regardless of sources and levels of the polygenic effects. The other three strategies lose power when the proportion of h_G^2 increases (Figure 4) or h_N^2 increases for a fixed h_G^2 (Supplementary Figure 3).

The BO strategy (with PERM_{F-M}) yields similar power to the ML strategy (with PERM_M and PERM_{F-M}) when h_G^2 remains constant at 15% (Supplementary Figure 3). When $h_G^2 \leq 5\%$, the BO strategy (with PERM_{F-M}) is more powerful than the ML strategy (with both permutation methods). More precisely, the BO strategy has about 12.3%, 10.7%, and 3.0% more power than the ML strategy for $h_G^2 = 0\%$, 5%, and 15%, respectively (Figure 4). As h_G^2 increases ($> 15\%$), the power for the ML and BO strategies becomes equivalent (Figure 4).

The ML and BO (with PERM_{F-M}) strategy are always more powerful than the UI and MO strategies (Figure 4 and Supplementary Figure 3) because the ML, BO strategies utilize the full sample and the UI and MO strategies include a partial sample.

Sample Size and Power

Power is estimated for varying sample sizes using simulation data for $h_G^2 = 15\%$ and $h_N^2 = 35\%$ at $\alpha = 0.001$. With a fixed h_g^2 level, the power is similar at $\text{MGF} = 2.5\%$ and 5% if other parameters remain the same. In the main text, we present power estimated at $\text{MGF} = 5\%$ for the UI, MO and ML strategies for varying sample sizes. The UI uses unrelated individuals, and the MO strategy uses only mother-offspring in the simulated data; whereas, the ML strategy uses the full sample. Therefore, we display the power for these three strategies in separate tables: Table 3 for the ML strategy, Supplementary Tables 1 and 2 for the UI and MO strategies, respectively.

Much smaller sample sizes are needed to detect a mitochondrial variant of $h_g^2 = 1\%$ with 80% or 90% power (Table 3, Supplementary Tables 1 and 2). Table 3 displays power estimations

for the ML strategy when $h_g^2=0.5\%$ and 1% . For family structure option I (nuclear family of size five), to detect a mitochondrial variant of $h_g^2=0.5\%$ with 80% or 90% power requires ~ 6,000 or 7,300 individuals, respectively (that is, 1,200 or 1,460 such nuclear families, respectively). For family structure option II (mixture of singletons and different family sizes), to detect a mitochondrial variant of the same effect size with 80% or 90% power requires ~ 5,400 (i.e. 729 singletons, 675 trios, 270 family size of five, and 162 family size of eight) or 6,200 (i.e. 837 singletons, 775 trios, 310 family size of five, and 186 family size of eight) individuals, respectively. With the same number of individuals, family structure II provides higher power than family structure I, because the latter structure includes more independent SSMLs than the former.

With unrelated individuals (i.e., founders in nuclear families or/and singletons in family structure II), we need ~ 3,400 and 4,400 subjects to reach 80% and 90% power, respectively, to detect a variant of $h_g^2=0.5\%$ (Supplementary Table 1). With mother-offspring data (for example, mother and 3 children), we need ~ 1,900 and ~ 2,200 nuclear families without fathers to reach 80% and 90% power to detect a variant of $h_g^2=0.5\%$, respectively (Supplementary Table 2). In considering cost-effectiveness, it is important to consider that only one individual in a SSML needs to be genotyped with respect to mitochondrial variants.

Application of the methods to FHS data

In this section, we first summarize the quality control procedures applied to the real data. Because the MO strategy yielded very similar results, we report association results based the ML strategy. In order to account for multiple testing, we performed permutation testing using the $PERM_M$ method to obtain empirical p-values for the ML strategy.

There were 65 examples of heteroplasmy among all genotyped mtSNPs in a total sample size of 2,455 individuals (heteroplasmy rate = $65/476,960 = 0.0001$). These heteroplasmy results were set as missing values. Among the 2,455 genotyped individuals, 1,286 belonged to 492 SSMLs, and 1,169 individuals were singletons. We compared SNPs among the same lineage members in the 492 SSMLs. Six genotypes were not consistent with other family members. We set these six genotypes to missing.

After inferring genotypes for non-genotyped individuals, the data set included 7,987 participants from three generations, belonging to 763 extended pedigrees. Supplementary Figure 4 displays $-\log_{10}(p\text{-values})$ for all phenotypes. The smallest p-value (p-value = 0.001) was obtained from the association between DBP at baseline with a mitochondrial variant at base pair position 10,590 (G10590A). This variant (MGF = ~1%) is a synonymous mutation that is located in the gene that encodes the enzyme mitochondrially encoded NADH dehydrogenase 4L (*MT-ND4L*). The empirical p-value for this association is 0.09, which is the smallest empirical p-value from the permutation test. This variant was previously found to have no association with colorectal cancer risk [Webb, et al. 2008].

Discussion

Owing to the unique mode of inheritance of the mitochondrial genome and the difficulty in separating the effects of the nuclear and mitochondrial genomes, different genotyping strategies, data cleaning procedures, and statistical analysis methods should be applied in association testing of mtSNPs using pedigree data. In this study, we performed an extensive simulation study to evaluate several analytical strategies to be used to account for potential polygenic effect(s) from the mitochondrial and/or nuclear genome for association testing of mitochondrial genetic variants using pedigree data. To evaluate empirical power, we compared two permutation methods. In addition, we estimated power under a series of effect sizes and sample sizes, which may help investigators plan studies of mtDNA using pedigree data. Furthermore, we discuss strategies for study design, genotyping, and data cleaning in association testing of mtDNA in pedigrees. Finally, we applied two of the analytical strategies to FHS data.

Permutation testing is used to obtain empirical p-values and to evaluate power for analytical strategies. In addition, it is used to account for multiple testing when many SNPs/traits are involved in association studies. To preserve heritability in pedigree data, a few permutation methods have previously been proposed in linkage analysis or association testing [Iturria, et al. 1999; MacCluer, et al. 1986; Wan, et al. 1997]. Nevertheless, these proposed permutation tests were only suitable for association testing of nuclear DNA with phenotypes. In this study, we explored two permutation methods, $PERM_{F-M}$ and $PERM_M$. For a study with only nuclear families, $PERM_{F-M}$ is likely to preserve most of the heritability (either from nDNA or mDNA) of a phenotype. However, $PERM_{F-M}$ may not be applicable in a prospective, long-term follow-up study consisting of both singletons and multiple generations of extended pedigrees (e.g. the FHS). $PERM_M$, utilizing the property of maternal inheritance, shuffles genotypes of a variant across SSMLs, which preserves relationship among mtSNPs in the same haplogroup (or phylogeny). This permutation method breaks the familial relationship and fails to preserve the nuclear heritability, thus, it yields very conservative empirical type I error rates for the NF and BO strategies, which take into account familial coefficients in model parameterization. Despite this disadvantage, $PERM_M$ did not require SSMLs to be the same size and was easy to perform. In addition, it yielded correct empirical type I error rates for the other three strategies.

Based on evaluations using both $PERM_{F-M}$ and $PERM_M$, the ML strategy was most robust in terms of empirical type I error and power among the three analytical strategies that used the full sample. For a study sample that contains only nuclear families, ML displays slightly lower power than BO only when $h_G^2 < 15\%$. If a study sample contains unrelated individuals along with extended pedigrees, ML is the best strategy to choose in combination with $PERM_M$.

We applied the ML strategy to 107 mitochondrial variants that were genotyped in the FHS. None of the association tests was significant after permutation analysis. These variants were identified through next generation sequencing using population data and may not be causal variants for metabolic traits. In another study, we reported the association with BP and/or FBG of mtDNA variants identified through sequencing the entire mitochondrial genome in

32 unrelated individuals with extreme-high BP in FHS and subsequently genotyped in 7000 individuals [Liu, et al. 2012].

Defining SSMLs is an essential first step in association testing of mtDNA and is useful in all phases of an association testing of mtDNA in pedigree data. We discuss several aspects pertaining to the use of SSML information in the Online Supplement.

In summary, a disease model that can better reflect the underlying biology can improve the utility of association studies. In this manuscript, we studied a disease model that takes into account polygenic effects from both the mitochondrial and nuclear genomes in analyzing pedigree data. We performed an extensive simulation study to evaluate power for different strategies in association testing of mtDNA. Our results provide guidance for association testing of mitochondrial variants in pedigrees.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NIH contract NO1-HC-25195 (to the Framingham Heart Study). The Center for Population Studies is supported by the Division of Intramural Research of the National Heart, Lung, and Blood Institute (Daniel Levy). We thank Dr. Kathy Lunetta of Biostatistics, School of Public Health at Boston University for discussions and providing R codes for implanting the BO strategy. We acknowledge Illumina, Inc for free genotyping of 2,455 FHS individuals with the *HumanOmni5*-Quad BeadChip. Illumina did not participate or had any influence on the analysis, interpretation, or reporting of the results.

References

- Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet.* 1994; 54(3):535–43. [PubMed: 8116623]
- Benn M, Schwartz M, Nordestgaard BG, Tybjaerg-Hansen A. Mitochondrial haplogroups: ischemic cardiovascular disease, other diseases, mortality, and longevity in the general population. *Circulation.* 2008; 117(19):2492–501. [PubMed: 18458168]
- Byrne EM, McRae AF, Duffy DL, Zhao ZZ, Martin NG, Whitfield JB, Visscher PM, Montgomery GW. Family-based mitochondrial association study of traits related to type 2 diabetes and the metabolic syndrome in adolescents. *Diabetologia.* 2009; 52(11):2359–68. [PubMed: 19760390]
- Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci.* 1963; 107:539–56. [PubMed: 14025561]
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* 2010; 42(2):105–16. [PubMed: 20081858]
- Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Analysis of European mtDNAs for recombination. *Am J Hum Genet.* 2001; 68(1):145–153. [PubMed: 11115380]
- Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* 2011; 12(6):R59. [PubMed: 21699709]
- Green DR. Apoptotic pathways: the roads to ruin. *Cell.* 1998; 94(6):695–8. [PubMed: 9753316]
- Hajnoczky G, Csordas G, Das S, Garcia-Perez C, Saotome M, Sinha Roy S, Yi M. Mitochondrial calcium signalling and cell death: approaches for assessing the role of mitochondrial Ca²⁺ uptake in apoptosis. *Cell Calcium.* 2006; 40(5–6):553–60. [PubMed: 17074387]

- He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*. 2010; 464(7288):610–4. [PubMed: 20200521]
- Herrnstadt C, Howell N. An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion*. 2004; 4(5–6):791–8. [PubMed: 16120433]
- Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*. 1988; 331(6158):717–9. [PubMed: 2830540]
- Iturria SJ, Williams JT, Almasy L, Dyer TD, Blangero J. An empirical test of the significance of an observed quantitative trait locus effect that preserves additive genetic variation. *Genet Epidemiol*. 1999; 17(Suppl 1):S169–73. [PubMed: 10597431]
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979; 110(3):281–90. [PubMed: 474565]
- Kim JH, Park KS, Cho YM, Kang BS, Kim SK, Jeon HJ, Kim SY, Lee HK. The prevalence of the mitochondrial DNA 16189 variant in non-diabetic Korean adults and its association with higher fasting glucose and body mass index. *Diabet Med*. 2002; 19(8):681–4. [PubMed: 12147150]
- Lal S, Madhavan M, Heng CK. The association of mitochondrial DNA 5178 C > a polymorphism with plasma lipid levels among three ethnic groups. *Ann Hum Genet*. 2005; 69(Pt 6):639–44. [PubMed: 16266403]
- Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavvas H, Cupples LA, Myers RH. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension*. 2000; 36(4):477–83. [PubMed: 11040222]
- Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009; 41(6):677–87. [PubMed: 19430479]
- Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet*. 2010; 87(2):237–49. [PubMed: 20696290]
- Li R, Liu Y, Li Z, Yang L, Wang S, Guan MX. Failures in mitochondrial tRNAMet and tRNAGln metabolism caused by the novel 4401A>G mutation are involved in essential hypertension in a Han Chinese Family. *Hypertension*. 2009; 54(2):329–37. [PubMed: 19546379]
- Liu C, Yang Q, Hwang SJ, Sun F, Johnson AD, Shirihai OS, Vasan RS, Levy D, Schwartz F. Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. *Hypertension*. 2012; 60(4):949–56. [PubMed: 22949535]
- Liu Y, Li R, Li Z, Wang XJ, Yang L, Wang S, Guan MX. Mitochondrial transfer RNAMet 4435A>G mutation is associated with maternally inherited hypertension in a Chinese pedigree. *Hypertension*. 2009; 53(6):1083–90. [PubMed: 19398658]
- MacCluer JW, VanderBerg JL, Read B, Ryder OA. Pedigree analysis by computer simulation. *Zoo Biol*. 1986; 5 (2):147–160.
- Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008; 456(7218):18–21. [PubMed: 18987709]
- McBride HM, Neuspiel M, Wasiak S. Mitochondria: more than just a powerhouse. *Curr Biol*. 2006; 16(14):R551–60. [PubMed: 16860735]
- Oh-hama T. Evolutionary consideration on 5-aminolevulinate synthase in nature. *Orig Life Evol Biosph*. 1997; 27(4):405–12. [PubMed: 9249985]
- Okura T, Koda M, Ando F, Niino N, Tanaka M, Shimokata H. Association of the mitochondrial DNA 15497G/A polymorphism with obesity in a middle-aged and elderly Japanese population. *Hum Genet*. 2003; 113(5):432–6. [PubMed: 12905068]
- Petersen KF, Befroy D, Dufour S, Dziura J, Ariyan C, Rothman DL, DiPietro L, Cline GW, Shulman GI. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science*. 2003; 300(5622):1140–2. [PubMed: 12750520]

- Poulton J, Luan J, Macaulay V, Hennings S, Mitchell J, Wareham NJ. Type 2 diabetes is associated with a common mitochondrial variant: evidence from a population-based case-control study. *Hum Mol Genet.* 2002; 11(13):1581–3. [PubMed: 12045211]
- Rossier MF. T channels and steroid biosynthesis: in search of a link with mitochondria. *Cell Calcium.* 2006; 40(2):155–64. [PubMed: 16759697]
- Samuels DC, Carothers AD, Horton R, Chinnery PF. The power to detect disease associations with mitochondrial DNA haplogroups. *Am J Hum Genet.* 2006; 78(4):713–20. [PubMed: 16532401]
- Saxena R, de Bakker PI, Singer K, Mootha V, Burt N, Hirschhorn JN, Gaudet D, Isomaa B, Daly MJ, Groop L, et al. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet.* 2006; 79(1):54–61. [PubMed: 16773565]
- Schork NJ, Guo SW. Pedigree models for complex human traits involving the mitochondrial genome. *Am J Hum Genet.* 1993; 53(6):1320–37. [PubMed: 8250048]
- Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM, Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. *Hum Mol Genet.* 2011; 20(8):1653–9. [PubMed: 21296868]
- Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB Sr, Fox CS, Larson MG, Murabito JM, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* 2007; 165(11):1328–35. [PubMed: 17372189]
- Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. *Nat Rev Genet.* 2005; 6(5):389–402. [PubMed: 15861210]
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. Classification of European mtDNAs from an analysis of three European populations. *Genetics.* 1996; 144(4):1835–50. [PubMed: 8978068]
- Torroni A, Lott MT, Cabell MF, Chen YS, Laverne L, Wallace DC. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet.* 1994; 55(4):760–76. [PubMed: 7942855]
- van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, Welsh-Bohmer KA, Saunders AM, Roses AD, Small GW, et al. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett.* 2004; 365(1):28–32. [PubMed: 15234467]
- Voet, D.; Voet, J.; Pratt, C. *Fundamentals of Biochemistry.* 2. John Wiley and Sons, Inc; 2006. p. 547
- Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ 2nd, Nikoskelainen EK. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science.* 1988; 242(4884):1427–30. [PubMed: 3201231]
- Wan Y, Cohen J, Guerra R. A permutation test for the robust sib-pair linkage method. *Ann Hum Genet.* 1997; 61(Pt 1):79–87. [PubMed: 9066930]
- Wang S, Li R, Fettermann A, Li Z, Qian Y, Liu Y, Wang X, Zhou A, Mo JQ, Yang L, et al. Maternally inherited essential hypertension is associated with the novel 4263 A>G mutation in the mitochondrial tRNA^{Ala} gene in a large Han Chinese family. *Circ Res.* 108(7):862–70. [PubMed: 21454794]
- Wang S, Li R, Fettermann A, Li Z, Qian Y, Liu Y, Wang X, Zhou A, Mo JQ, Yang L, et al. Maternally inherited essential hypertension is associated with the novel 4263A>G mutation in the mitochondrial tRNA^{Ala} gene in a large Han Chinese family. *Circ Res.* 2011; 108(7):862–70. [PubMed: 21454794]
- Webb E, Broderick P, Chandler I, Lubbe S, Penegar S, Tomlinson IP, Houlston RS. Comprehensive analysis of common mitochondrial DNA variants and colorectal cancer risk. *Br J Cancer.* 2008; 99(12):2088–93. [PubMed: 19050702]
- Wilson FH, Hariri A, Farhi A, Zhao H, Petersen KF, Toka HR, Nelson-Williams C, Raja KM, Kashgarian M, Shulman GI, et al. A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. *Science.* 2004; 306(5699):1190–4. [PubMed: 15498972]

A. Coefficient structure for nDNA

$$\Phi = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.5 \\ 0 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

B. Coefficient structure for mtDNA

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Figure 1.

Covariance structures for polygenic effects from nuclear and mitochondrial genomes, respectively, in a nuclear family of size five (two parents and three offspring). nDNA = nuclear DNA, mtDNA = mitochondrial DNA.

Figure 2.

Type I error rates estimated at $\alpha = 0.001$ when $h_g^2 = 0\%$ for $h_G^2 + h_N^2 = 50\%$. The y-axis is plotted on a log10 scale, therefore, 0.001, 0.002, 0.003, 0.004, 0.005 correspond to -3.0 , -2.70 , -2.52 , -2.40 , and -2.30 on the log10 scale, respectively. The x-axis is $h_g^2(\%)$. The five strategies considered were as follows: 1) UI - unrelated individuals (i.e., founders or singletons) were analyzed using simple linear regression. 2) NF - all individuals were included and familial correlation (Figure 1A) due to autosomal polygenic component among nuclear families was accounted for in a linear mixed effects (LME) model. 3) MO - fathers were excluded from the sample, but mother and offspring were used for analysis. This reduced covariance structure (without father in Figure 1A or 1B) could be accommodated by LME. 4) ML - all individuals were used. Random effect in SSML (Figure 1B) was accounted for in a LME model. 5) BO - all individuals were used. Both coefficient structures of nDNA (Figure 1A) and mtDNA (Figure 1B) were accounted for in a LME model.

Figure 3.

Empirical type I error rates corresponding to $\alpha = 0.001$ for the two permutation tests when $h_G^2 + h_N^2 = 50\%$. The y-axis is plotted on a log10 scale. Therefore, 0.001, 0.0001, and 0.00001 correspond to -3.0 , -4 , and -5 on the log10 scale, respectively. The x-axis is $h_g^2(\%)$. Please see footnote in Figure 2 for description of the analytical strategies. In the parentheses of the legend, “M” stands for maternal permutation or PERM_M and “F-M” stands for father-mother pair permutation or PERM_{F-M} .

Figure 4.

Power estimated using $\alpha_{PERM_F_M_0.001}$ and $\alpha_{PERM_M_0.001}$ when $h_G^2 + h_N^2 = 50\%$. Please see footnote in Figure 2 for description of the analytical strategies. In the parenthesis of the legend, “M” stands for maternal permutation or $PERM_M$ and “F-M” stands for father-mother pair permutation or $PERM_{F-M}$.

Table 1

General settings for the unknown polygenic effects from nDNA and mtDNA

A. $h_G^2 + h_N^2 = 50\%$						
h_G^2 (%)	0	5	15	25	35	45
h_N^2 (%)	50	45	35	25	15	5

B. $h_G^2 = 15\%$				
h_G^2 (%)	15	15	15	15
h_N^2 (%)	0	5	15	35

Table 2

Sample size and family structure in the secondary simulation analysis

Total individuals		2000	3000	4000	5000	6000	7000	8000	9000	10000
I	Parents + 3 kids	400	600	800	1000	1200	1400	1600	1800	2000
II										
	Singleton	270	405	540	675	810	945	1080	1215	1350
	Parents + 1 kids	250	375	500	625	750	875	1000	1125	1250
	Parents + 3 kids	100	150	200	250	300	350	400	450	500
	Parents + 6 kids	60	90	120	150	180	210	240	270	300

Table 3

Power estimated at $\alpha_{PERM_M_0.001}$ for the ML strategy when MGF = 5%, $h_g^2=0.5\%$, $h_G^2=15\%$ and $h_N^2=35\%$.

Number of individuals	Power (%)			
	Nuclear family with 3 offspring ^a		Varying number of family size ^b	
	$h_g^2=0.5\%$	$h_g^2=1\%$	$h_g^2=0.5\%$	$h_g^2=1\%$
2000	20.1	47.8	25.2	64.0
3000	36.4	65.5	46.4	87.5
4000	55.5	81.4	64.3	96.6
5000	69.3	86.2	78.3	99.1
6000	79.9	89.6	88.1	99.9
7000	88.0	93.5	93.7	99.9
8000	93.1	95.8	97.0	100.0
9000	96.1	97.5	98.5	100.0
10000	98.2	98.9	99.3	100.0

^a Samples included nuclear families of equal sizes: parents with 3 offspring.

^b Mixed family sizes: samples included, singletons, trios, nuclear families of size five (two parents and three siblings) and nuclear families of size eight (two parents and six siblings) (ratios =27:75:50:48).