Correlation of Intergenerational Family Sizes Suggests a Genetic Component of Reproductive Fitness

Anna Pluzhnikov,* Daniel K. Nolan,*,† Zhiqiang Tan,‡ Mary Sara McPeek, and Carole Ober

Reproductive fitness is a complex phenotype that is a direct measure of Darwinian selection. Estimation of the genetic contribution to this phenotype in human populations is confounded by within-family correlations of sociocultural, economic, and other nongenetic factors that influence family sizes. Here, we report an intergenerational correlation in reproductive success in the Hutterites, a human population that is relatively homogeneous with respect to sociocultural factors that influence fertility. We introduce an estimator of this correlation that takes into account the presence of multiple parent-offspring pairs from the same nuclear family. Statistical significance of the estimated correlation is assessed by a permutation test that maintains the overall structure of the pedigree. Further, temporal trends in fertility within this population are accounted for. Applying these methods to the S-Leut Hutterites yields a correlation in effective family size of 0.29 between couples and their sons and 0.18 between couples and their daughters, with empirical $P < 1 \times 10^{-6}$ and $P = .0041$, respectively. Similar results were obtained for completed families (0.31 between couples and their sons and 0.23 between couples and their daughters; empirical $P < 1 \times 10^{-6}$ and $P = .00059$, respectively). We interpret these results as indicating a significant genetic component to reproductive fitness in the Hutterites.

Reproductive fitness is the relative proportion of genes that one contributes to subsequent generations. Formally, it is a measure of both fertility (the ability to reproduce) and mortality (the survival of children to reproductive age). Both components of fitness and the ability to reproduce and to have children that also reproduce are complex phenotypes, influenced by genetic and environmental factors. Studies of model organisms suggest that hundreds of genes directly influence fertility,[1,2] and thousands of human genetic disorders that influence the ability to reproduce have been described (OMIM). Although it is relatively straightforward to study the second component of fitness in humans, fertility potential is difficult to assess, because human family sizes are often deliberately limited for a variety of reasons, and few populations reach their true reproductive potential, because of sociocultural or economic reasons. Thus, although a handful of genes have been identified that cause infertility in humans (reviewed by Layman[3]), it has been difficult to assess genetic contributions to fertility. For example, parent-child correlations in family sizes have been reported in many populations (table 1); however, in each study, the contribution of nongenetic factors to the correlation could not be excluded, and, in all cases, the correlation was attributed to social or cultural phenomena.[4–9] Thus, the relative contribution of genes to fertility differences between couples in human populations is currently not known.

The Hutterites are a founder population of European descent that offers the opportunity to study human fertility potential.[10,11] The Hutterites traditionally proscribe contraception, and, as a result, median completed sibship sizes among the S-Leut Hutterites are large (median sibship size was >10 in the 1960s[12]), and interbirth intervals are small (mean interval was <2 years between births through the 10th birth during this same time period[10]). Moreover, the Hutterite communal lifestyle assures that sociocultural factors known to influence family size, including wealth, access to medical care, age at marriage, desire for large families, and nursing practices, are remarkably uniform in the population.[13] The naturally high fertility rate coupled with their communal lifestyle makes them ideally suited for genetic studies of fertility, as noted elsewhere.[10,11]

As a first step in assessing genetic contributions to fertility, we sought to determine whether intergenerational correlations in family size were present in the Hutterites, as they are in other populations.[4–9] However, because of the general absence of variation in nongenetic factors that are known to influence family sizes, parent-children correlations in Hutterite family sizes would indicate a genetic component (i.e., heritability) to this complex trait. Two measures of family size have been used in previous studies. The first, census family size (CFS), is defined as the total number of children born to a couple, whereas the second, effective family size (EFS), is defined as the number of children who reproduce in the population per reproducing individual.[6] Differences in CFS and EFS could reflect

Table 1. Correlations in Family Size in Different Populations

Population and Measure of Family Size	Correlation in Family Size		
	Couple-Children	Father-Sons	Mother-Daughters
British[4]:			
CFS	NR	.066[a]	.21[a]
Japanese[5]:			
CFS	NR	−.031	.014
French Canadians[6]:			
EFS	.16	.14	.17
Icelanders[7]:			
CFS	NR	.041	.079
Irish[8]:			
CFS	NR	NR	~.17[b]
Hutterites:			
r_{reg}:			
EFS	NR	.29[c]	.18[d]
CFS	NR	.31[c]	.23[d]
Pearson's r:			
EFS	NR	.34[c]	.19[d]
CFS	NR	.34[c]	.24[d]

NOTE.—All correlations are estimated by Pearson's r except where noted. NR = not reported.

[a] Correlation not confirmed in a subsequent study that accounted for time trend in data.

[b] Estimated from reported heritability of .34.

[c] Couple-son correlation.

[d] Couple-daughter correlation.

differences in out-migration or marriage rates between families or in children's mortality rates prior to reproductive age. Whereas the former is not likely to be genetic in origin, the latter could include genetic causes.

For these studies, we used demographic data for the S-Leut Hutterites that are part of the database originally collected by Dr. A. G. Steinberg nearly 50 years ago[11] and maintained and updated at the University of Chicago by C.O.[14] For this study, the 13,165-person S-Leut genealogy was divided into 1,037 3-generation subpedigrees. We constructed 3-generation subpedigrees, beginning with the founders, and moved down through the pedigree (a top-down approach), in contrast to a coalescent (or bottom-up) approach in which pedigrees are constructed backward in time from current descendants, as was used for the older periods (18th century) in one study.[6] In the coalescent approach, only ancestors with descendants in the current population are represented in the sample; therefore, no couples with EFS values of zero would be included. With the top-down approach, couples with an EFS of zero are included as long as they do not prevent the formation of a 3-generation pedigree. Therefore, couples with EFS of zero would not be included in the grandparental (top) generation but would be included in the second (middle) generation if at least one of the children in that generation has a nonzero EFS. In the third (bottom) generation, it is possible for all values to be zero.

To ensure that only completed families were included in our sample, we considered 3-generation pedigrees in which the grandmother (top generation) was born in or before 1860. We assumed that nearly all women born in 1940 or earlier would have completed their families by 1980 (the date of the most recent complete update of our database), since relatively few Hutterite women have children after age 40 years. Thus, by setting the upper limit at 1860, we ensure that the children in the 3rd generation would have reached age 40 years and likely completed their families by 1980. Of the 1,037 subpedigrees, 235 met this criterion. We excluded an additional 75 couples because one spouse from each couple died before reaching age 40 years. In the remaining 160 3-generation pedigrees, one male had a marriage and remarriage, both of which met the above criteria and were therefore included as two separate families in our analysis. Our final sample included 161 3-generation pedigrees.

Previous studies of CFS or EFS used the standard estimator of correlation (Pearson's r), which assumes that the data are in the form of independent pairs (consisting of a couple's EFS or CFS and an offspring's EFS or CFS). However, for each Hutterite couple, we generally have information about all couples' reproducing offsprings' EFS or CFS values, rather than for a single reproducing offspring. Therefore, to allow for the fact that a couple's EFS or CFS would be compared with multiple offsprings' EFSs or CFSs, we developed a regression-based estimator of the correlation in EFS and CFS, which we refer to as "r_{reg}." Our estimator of correlation was derived as follows. For $j = 1, \ldots, m$, where m is the number of 3-generation pedigrees, let M_j be the jth couple's EFS or CFS—that is, the number of reproducing children. Given M_j, for $i = 1, \ldots, M_j$, let Y_{ji} be the EFS or CFS of the ith reproducing child of the jth couple. We assume that $E(M_j) = \mu_M$, $\text{Var}(M_j) = \sigma_M^2$, $E(Y_{ji}|M_j) = \alpha + \beta M_j$, and $\text{Var}(Y_{ji}|M_j) = \sigma_Y^2$. The ordinary least-squares estimators of the parameters, obtained from simple linear regression of Y on M (see, e.g., chapter 1 in the work of Weisberg[15]), would then be

$$\hat{\mu}_M = \bar{M} \ ,$$

$$\hat{\sigma}_M^2 = \frac{1}{m-1}\sum_{j=1}^{m}(M_j - \bar{M})^2 \ ,$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\tilde{M} \ ,$$

$$\hat{\beta} = \frac{\sum\limits_{j=1}^{m}\sum\limits_{i=1}^{M_j}(M_j - \tilde{M})(Y_{ji} - \bar{Y})}{\sum\limits_{k=1}^{m} M_k(M_k - \tilde{M})^2} \ ,$$

and

$$\hat{\sigma}_Y^2 = \frac{1}{m\bar{M} - 2}\sum_{j=1}^{m}\sum_{i=1}^{M_j}(Y_{ji} - \hat{\alpha} - \hat{\beta}M_j)^2 \ ,$$

where

$$\bar{M} = \frac{1}{m}\sum_{j=1}^{m} M_j \ ,$$

$$\tilde{M} = \frac{\sum_{j=1}^{m} M_j^2}{\sum_{k=1}^{m} M_k} \ ,$$

and

$$\bar{Y} = \frac{\sum_{j=1}^{m}\sum_{i=1}^{M_j} Y_{ji}}{\sum_{k=1}^{m} M_k} \ .$$

(We actually use m and $m\bar{M}$ in the denominators of $\hat{\sigma}_M^2$ and $\hat{\sigma}_Y^2$, respectively, which give asymptotically equivalent estimators.) This leads to the following estimator of the correlation, r_{reg}, between M and Y

$$\hat{r}_{\mathrm{reg}} = \frac{\hat{\beta}}{\sqrt{\hat{\beta}^2 + \dfrac{\hat{\sigma}_Y^2}{\hat{\sigma}_M^2}}} \ .$$

Although we are not aware of any sociocultural factors that could influence interindividual differences in Hutterite family sizes, it remained possible that a time trend in the data, or temporal heterogeneity, was present that could affect a correlation in family size.[7] To explore this possibility, we examined the relationship between family size and generation number for the 161 pedigrees. Because the S-Leut pedigree contains overlapping generations and the founders (ordinarily generation 0) entered at different time points over ~100 years, we scored generations in a way that related generation number to absolute time by first assigning the founder couples born between 1700 and 1800 to generation 0. We then incremented the generation number of their children by one. When we assigned generation number to children of nonfounders, the highest generation of either parent was incremented by one. When a founder married a nonfounder, we used the nonfounder generation to determine the generation number of their children. Because ages of Hutterite spouses are highly correlated ($r = 0.955$), this latter step aligned generation number with absolute time. EFS values plotted by generation revealed temporal heterogeneity (fig. 1). The two most recent generations (6 and 7) had higher mean EFS values than did generations 0–5, which were all similar. Interestingly, generations 6 and 7 correspond to the period following the Hutterite migration to North America in the late 1870s, which was associated with an exponential increase in population size.[13] We corrected for temporal trends by subtracting the mean EFS (or CFS) value for each generation from the EFS (or CFS) value for each individual within that same generation. This created an
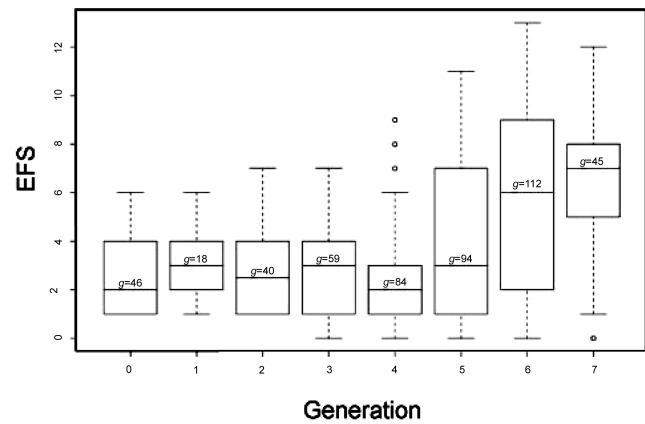

Figure 1. EFS values by generation. Data are shown for the 161 3-generation subpedigrees, where g is the number of individuals within each generation and horizontal bars indicate the variance. Outliers are shown as unblackened circles, and the mean EFS is represented by a horizontal line within each rectangle.

approximately normal distribution of EFS and CFS values, centered at zero. These "corrected" values were used in our analyses. With use of the regression-based estimator, the correlation between the EFS of couples and the EFS of their sons was $r_{\mathrm{reg}} = 0.29$ and between the EFS of couples and the EFS of their daughters was $r_{\mathrm{reg}} = 0.18$, which were high compared with estimates of correlations in family sizes in other populations (table 1 also shows the uncorrected values for comparison with previous studies). The correlations in the EFS values with sons and with daughters are illustrated in figure 2.

To assess the significance of these correlations, we used a permutation test that both accounted for the lack of independence among the couple-child pairs and maintained the overall structure of the pedigree. To maintain the shape of the pedigrees, individuals in the 3-generation subpedigrees were separated by their generation number, and EFS and CFS values were randomly permuted among couples within each generation (i.e., horizontally in the pedigree). This process was repeated for the entire pedigree 1,000,000 times, and, each time, the estimator of correlation, r_{reg}, was determined for correlations with sons and for correlations with daughters. The value of r_{reg} from the data was then compared with the null distribution of the correlations from 1,000,000 repeated permutation processes, to obtain an empirical P value. This strategy of permuting only the EFS (or CFS) values of contemporaneous individuals has the added benefit of maintaining any time trends that may exist in the data that are not otherwise accounted for. With use of this approach, the likelihood of observing a correlation in EFS between a couple and their sons of at least 0.29 is $<1 \times 10^{-6}$ and of observing a correlation in EFS between a couple and their daughters of at least 0.18 is .0041. The corresponding values for CFS were 0.31 ($P < 1 \times 10^{-6}$) for sons and 0.23
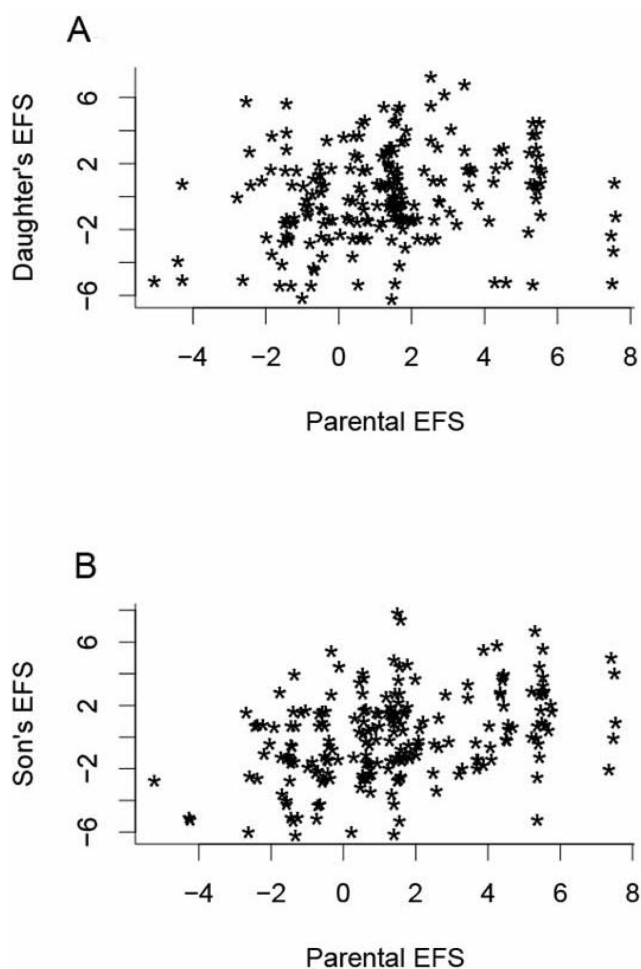
Figure 2. Scatterplots showing correlation in EFS values for couples and sons (*A*) and couples and daughters (*B*). EFS values are adjusted for generation.

($P = .00059$) for daughters. These highly significant correlations in parent-child EFS and CFS suggest a heritable (i.e., genetic) component to reproductive fitness in the S-Leut Hutterites.

Only one previous study, of the population of Saguenay-Lac Saint Jean (Quebec), considered correlations in EFS between generations.[6] Although EFS was correlated between generations (table 1), CFS was not. Those authors interpreted this difference as reflecting the greater outmigration among some families and concluded, therefore, that the correlations in EFS were due to sociocultural factors. In the Hutterites, correlations in CFS and EFS were both significant, further supporting the role of genetic differences between families that underlie these correlations. Other nongenetic factors that have explained correlations in CFS are social transmission of reproductive behavior in Iceland,[7] differential access to resources (including ability to acquire a mate) in western Ireland,[8] socioeconomic class in 19th-century England,[4] the introduction of birth-control use in Japan,[5] wealth (most importantly, its effect on

number of wives) in South American Indian tribes,[9] and cultural factors (such as cooperation based on kinship) in hunter-gather populations.[16]

In these regards, the Hutterites are remarkable in a number of ways. First, they are strictly monogamous, and only 2% of Hutterite couples are childless, compared with 10% of the general population.[17] Second, there is a very small variance in the wife's age at marriage, with most women marrying at age 20–24 years and few (~10%) after age 24 years.[10] Third, their communal lifestyle guarantees equal access to education, income, food, health care, and housing.[13] Fourth, their religion proscribes contraception; even today, birth-control use is minimal and family sizes are large.[18] Lastly, the lengths of time that women nurse their infants are relatively uniform among Hutterites, although nursing length for daughters is longer than that for sons.[19]

Because of the absence of any known sociocultural factor that could influence differences in family sizes among Hutterite families, we interpret the results of our study as indicating that genetic differences among couples influence reproductive fitness in this population. This conclusion is supported by previous studies of the Hutterites that showed decreased fertility among the most inbred women[20] and genomewide transmission distortion among sibships.[21] Those studies suggested that genetic variation segregating in the population influences reproductive success or fetal survival, respectively. Because most genetic variation that is present in the Hutterites is also present in outbred European populations,[22,23] it is likely that common variation influencing fertility in the Hutterites is also present in outbred populations. On the other hand, the reduced environmental heterogeneity may enhance the effects of genetic variation on complex phenotypes in this population, making them easier to detect in the Hutterites compared with more-heterogeneous populations, as discussed elsewhere.[24]

To our knowledge, this is the first report of correlations in human family size that cannot be explained by sociocultural or other nongenetic differences between families. Although we cannot exclude the possibility that unknown differences exist, we think it most likely that the observed differences reflect genetic differences between couples. In this context, two observations are noteworthy. First, there is stronger correlation for CFS than for EFS between couples and their daughters, but that is not the case between couples and their sons. This likely reflects the fact that, because of the excess number of adult women compared with adult men, nearly all Hutterite men but not all Hutterite women marry. Because the probability of marriage (and subsequent reproduction) is not genetic, there would be more nongenetic variation influencing EFS, compared with CFS, among Hutterite women. Second, there is a stronger correlation in both EFS and CFS between couples and their sons than between couples and their daughters. This difference could reflect, at least in part, the influence of both Y-chromosome and mitochondrial genes on male reproduction, since mutations in both have been associ-

ated with male infertility.[25,26] However, this difference may also reflect the overall larger number of genes involved in male compared with female reproductive processes.[2] Nonetheless, the highly significant correlations in family sizes in the Hutterites suggest that genes influencing reproductive fitness are segregating in the population, making this complex phenotype amenable to mapping studies for the identification of human fertility genes.

Web Resource

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

References

1. Matzuk MM, Lamb DJ (2002) Genetic dissection of mammalian fertility pathways. Nat Cell Biol Suppl 4:s41–s49
2. Torgerson DG, Whitty BR, Singh RS (2005) Sex-specific functional specialization and the evolutionary rates of essential fertility genes. J Mol Evol 61:650–658
3. Layman LC (2002) Human gene mutations causing infertility. J Med Genet 39:153–161
4. Pearson KAL, Bramley-Moore L (1899) On the inheritance of fertility in mankind. Phil Trans Royal Soc Lond 192:257–230
5. Imaizumi Y, Nei M, Furusho T (1970) Variability and heritability of human fertility. Ann Hum Genet 33:251–259
6. Austerlitz F, Heyer E (1998) Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. Proc Natl Acad Sci USA 95:15140–15144
7. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefánsson K (2003) A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am J Hum Genet 72:1370–1388
8. Madrigal L, Relethford JH, Crawford MH (2003) Heritability and anthropometric influences on human fertility. Am J Hum Biol 15:16–22
9. Neel JV (1970) Lessons from a "primitive" people. Science 170:815–822
10. Sheps MC (1965) An analysis of reproductive patterns in an American isolate. Pop Stud 19:65–80
11. Steinberg AG, Bleibtreu HK, Kurczynski TW, Martin AO, Kurczynski EM (1967) Genetic studies in an inbred human isolate. In: Crow JF, Neel JV (eds) Proceedings of the Third International Congress of Human Genetics. Johns Hopkins University Press, Baltimore, pp 267–290
12. Mange AP (1964) Growth and inbreeding of a human isolate. Human Biol 36:104–133
13. Hostetler JA (1974) Hutterite society. Johns Hopkins University Press, Baltimore
14. Ober C, Cox NJ (1998) Mapping genes for complex traits in founder populations. Clin Exp Allergy Suppl 28:101–105
15. Weisberg S (1985) Applied linear regression. Wiley, New York
16. Blum MG, Heyer E, Francois O, Austerlitz F (2006) Matrilineal fertility inheritance detected in hunter-gatherer populations using the imbalance of gene genealogies. PLoS Genet 2:e122
17. Ober C, Hyslop T, Elias S, Weitkamp LR, Hauck WW (1998) Human leukocyte antigen matching and fetal loss: results of a 10-year prospective study. Hum Reprod 13:33–38
18. Ober C, Hyslop T, Hauck WW (1999) Inbreeding effects on fertility in humans: evidence for reproductive compensation. Am J Hum Genet 64:225–231
19. Margulis SW (1996) The effects of inbreeding on parental behavior and reproductive success in two species of the old-field mouse (*Peromyscus polionotus*). University of Chicago Press, Chicago
20. Ober C, Karrison T, Odem RR, Barnes RB, Branch DW, Stephenson MD, Baron B, Walker MA, Scott JR, Schreiber JR (1999) Mononuclear-cell immunisation in prevention of recurrent miscarriages: a randomised trial. Lancet 354:365–369
21. Zöllner S, Wen X, Hanchard NA, Herbert MA, Ober C, Pritchard JK (2004) Evidence for extensive transmission distortion in the human genome. Am J Hum Genet 74:62–72
22. Newman DL, Hoffjan S, Bourgain C, Abney M, Nicolae RI, Profits ET, Grow MA, Walker K, Steiner L, Parry R, et al (2004) Are common disease susceptibility alleles the same in outbred and founder populations? Eur J Hum Genet 12:584–590
23. Ober C, Cox N, Parry R, Abney M, DiRienzo A, Changyaleket B, Gidley H, Kurtz B, Lander ES, Lee J, et al (1998) Genome-wide search for asthma susceptibility loci in a founder population. Hum Mol Genet 7:1393–1398
24. Ober C, Abney M, McPeek MS (2001) The genetic dissection of complex traits in a founder population. Am J Hum Genet 69:1068–1079
25. McElreavey K, Ravel C, Chantot-Bastaraud S, Siffroi JP (2006) Y chromosome variants and male reproductive function. Int J Androl 29:298–303
26. St John JC, Jokhi RP, Barratt CL (2005) The impact of mitochondrial genetics on male infertility. Int J Androl 28:65–73