

Inheritance of Total Serum IgE in the Isolated Tangier Island Population from Virginia: Complexities Associated with Genealogical Depth of Pedigrees in Segregation Analyses

Rasika A. Mathias^{a,b} Terri H. Beaty^a Joan E. Bailey-Wilson^b Carol Bickel^c
Maria L. Stockton^c Kathleen C. Barnes^c

^aDepartment of Epidemiology, Bloomberg School of Hygiene and Public Health, Johns Hopkins University;

^bInherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health,

and ^cDivision of Allergy and Clinical Immunology, School of Medicine, Johns Hopkins University, Baltimore, Md., USA

Key Words

Isolate · Inbreeding · IgE · Segregation analysis

Abstract

Objectives: This study was aimed at performing a segregation analysis of total serum immunoglobulin E (tIgE) in an isolated population using maximal genealogical information permitted by current software and computer capacities, while assessing the reliability of the best-fitting model of inheritance for tIgE through simulations.

Methods: All current Tangier Island, VA, residents (n = 664) belonged to one large extended pedigree (n = 3,501) spanning 13 generations, with an average inbreeding coefficient of 0.009. Phenotype data were obtained on 453 (68.2%) of the residents using a population-based recruitment scheme. Due to computational limitations resulting from the extremely complex pedigree structure, analysis on only two pedigree reconstructions was feasible: a reduced pedigree retaining all phenotyped individuals and their parents as 57 distinct families, and 922 nuclear families. **Results:** Familial correlations and heritability calculations reveal a significant genetic component to tIgE in these data (heritability = 26%). The most

parsimonious model to explain tIgE distribution indicated by the reduced pedigree structure was a two-distribution Mendelian model. However, larger and non-genetic models could not be rejected. Simulations over 200 replicates performed to evaluate the reliability of this model, indicated that using restricted genealogical information had minimal impact on results of segregation analyses performed here.

Copyright © 2005 S. Karger AG, Basel

Introduction

Across the globe, rates of asthma prevalence have been steadily increasing over the past several decades [1, 2]. While it has been argued that many of these trends may be attributed to changes in research methods and diagnostic criteria, most researchers agree the observed increases are real. The prevalence of asthma varies greatly among countries, with higher prevalence observed in Western and developed countries compared to developing countries [3]. The vast majority of asthmatic individuals experience symptoms following exposure to an inhaled trigger, and in these individuals there appears to

be a strong relationship between asthma, allergen exposure, Immunoglobulin E (IgE) production, and atopy (allergic response associated with elevated IgE) [4, 5].

Several researchers have demonstrated higher concordance rates in total serum IgE (tIgE) among monozygotic twins compared to dizygotic twins [6, 7], with heritability estimates ranging between 60% [8] and 92% [9]. There have also been several segregation analyses of tIgE, and a variety of models have been proposed. These studies have provided evidence for recessive inheritance of 'high' IgE [10–13], polygenic [14], codominant [15–17], mixed models [18, 19], as well as dominant models for 'high' IgE [20, 21]. Conversely, a study by Parks et al. [22] found no evidence for a genetic control of tIgE, but suggested that parasitism which was endemic in their population could potentially confound the underlying IgE distribution. Clearly, the mode of inheritance of tIgE remains uncertain. One explanation may be genetic heterogeneity – both within any single study (i.e. between families) and across studies. The issues of ascertainment for highly correlated traits such as asthma may also contribute to inconsistent results across studies. Samples ascertained through asthmatic or allergic family members are more likely to have a bimodal distribution for IgE with asthmatic/allergic individuals having higher IgE levels, and these could easily push evidence in favor of a dominant or recessive model. Proper adjustment for appropriate covariates to remove background noise is another concern. In samples of families enriched for atopic individuals, the season of bleed becomes relevant when studying tIgE levels. The goal of this study was to perform a segregation analysis of tIgE using an isolated population with comprehensive genealogical information, including only members from a single large pedigree with relatively uniform allergen exposure.

Materials and Methods

Study Population

Tangier Island, Virginia, is a 3 by 1-mile wide marshy piece of land approximately 2 feet above sea level. It lies close to the mouth of the Chesapeake Bay, 12 miles from the nearest mainland port (Crisfield, Maryland). The current population is approximately 650, and the isolated nature of the island has led to low levels of inbreeding as described elsewhere [23]. Briefly, the average inbreeding coefficient (F) for the entire Tangier Island Population (dating back to 1722) is 0.009 (range: 0.000–0.156, $n = 3,512$). In the early cohorts close consanguinity contributed to inbreeding, while in the later cohorts there was a conscious avoidance of close matings and the observed inbreeding levels were a result of older lines of more distant relationships. The prevalence of asthma on Tangier Island is 5% ($n = 33$), and while up to 57% of the popula-

tion has self-reported rhinitis, only 19% have documented allergic rhinitis.

Subject Characterization

Volunteer subjects from Tangier Island were phenotyped to measure levels of tIgE. All subjects gave written informed consent as approved by The Johns Hopkins University Institutional Review Board. Subject characterization included: (1) an interviewer-administered standardized respiratory questionnaire from which data on age, gender and smoking exposure were obtained, and (2) peripheral fasting blood collection for IgE. All residents of Tangier Island were invited to participate in this study, regardless of their asthma or atopy status.

Total Serum IgE Levels

Measurements were conducted on serum samples using chemiluminometric immunoassay (Magic Lite Total IgE extended range; ALK, Copenhagen, Denmark/CIBA-Corning, Medfield, Mass., USA) in the Dermatology Allergy and Clinical Immunology laboratories at Johns Hopkins University, Baltimore. All measurements were conducted in duplicate and the average of the two values was used. For samples with coefficient of variation (CV) >15%, measurements were retested, and original values were discarded.

Preliminary Data Analysis

Several groups have argued that the presence of significant skewness in the distribution of a quantitative trait such as tIgE can lead to falsely inferring a major gene [24]. This is because skewness can mimic the commingling of two (or more) underlying distributions. Consequently, the first stage of analysis with these data was to test for normality in tIgE, and choose a suitable transformation. Unpaired t tests were used to compare mean levels of tIgE between dichotomous covariates. For categorical variables with k levels, linear regressions with $k - 1$ dummy variables were used. Simple linear regression was used to evaluate the relationship between normalized tIgE and age. Finally, multiple linear regression models were used to adjust for all covariates showing independent relationships with normalized tIgE. The intercept was added back to residual values to maintain the original scale.

Familial Correlations

Familial correlations for parent-offspring (also father-child and mother-child), spousal, and sib-sib (also broken down by gender) and their significance were calculated using Genetic Analysis Software (GAS v.2.0).

Pedigree Feasibility

A simple model of inheritance was fit using several pedigree reconstructions to determine run times for segregation analysis given a single set of starting values. In this testing phase, a simple polygenic model estimating one mean, standard deviation, and heritability was fit for various pedigree reconstructions. Testing was conducted on a high-speed SUN station with two 480 MHz processors and 2 gigabytes of memory. At each stage, the computation time taken for a single iteration in the maximum likelihood procedure was noted. Estimating on average 20 iterations to reach a maximum point on the likelihood surface, the estimated run time for a single set of initial values was calculated. Pedigree reconstructions attempted are described in table 1, and ranged from the full complex pedigree to 922 nuclear pedigrees with the duplication of 297 phenotyped individuals.

Table 1. Feasibility estimates of a segregation analysis with varying levels of pedigree structure in the Tangier Island population

Pedigree structure	CPU time for 1 iteration	Time per maximization	Number of pedigrees	Number of inbred subjects
Full 13-generational pedigree	>14 days	~9 months	1	1,917
Upper 5 generations removed	>7 days	~4.5 months	10	19
445 phenotyped individuals				
+ 1 generation above	~3 min	<1 h	57	0
445 phenotyped individuals				
+ 2 generations above	>7 days	~4.5 months	19	9
Nuclear families with duplication	<20 s	<4 min	922	0

Segregation Analysis

Segregation analyses were carried out using Pedigree Analysis Package (PAP v.4.0) [25]. Because of the population-based recruitment scheme, no ascertainment correction was necessary. A general unrestricted model with a single locus, three genotypic means, single variance, residual heritability, and estimated transmission probabilities was fit. Several restricted models nested within this general model were considered including: (1) a non-genetic single-distribution sporadic model; (2) a simple polygenic model; (3) *Mendelian dominant, codominant and recessive mixed* models, and (4) *environmental* models with equal transmission of a non-genetic factor controlling tIgE.

The Likelihood Ratio Test (LRT)

The likelihood of a model to explain the distribution of a trait in a pedigree is proportional to the probability of the observed phenotypic data under this model [26]. Once maximal log-likelihood values have been attained within PAP, nested models may be compared using the LRT. Asymptotically, if the reduced model is true, then twice the difference in the ln-likelihoods of the two models will follow a chi-square distribution with n degrees of freedom (d.f.); where n is the difference in the number of parameters estimated in the two models. However, when parameters are constrained at their boundaries in the reduced model, the LRT is more approximately compared to a mixture of χ^2 distributions, and in these situations mixtures of chi-squares as suggested by Self and Liang [27] can be considered. Briefly, as an example consider the simple case of a quantitative trait in the context of a mixed single locus model where the general model has nine estimated parameters: the allele frequency (q_A), three genotypic means (μ_{AA} , μ_{AB} , and μ_{BB}), common variance (σ^2), residual heritability (h^2), and three arbitrary transmission parameters (τ_{AA} , τ_{AB} , and τ_{BB}). The mixed codominant Mendelian model is nested within this general model, and has six estimated parameters: q_A , μ_{AA} , μ_{AB} , μ_{BB} , σ^2 , and h^2 , with the transmission probabilities (τ_{AA} , τ_{AB} , and τ_{BB}) constrained to their Mendelian expectations of 1, 0.5 and 0, respectively. In routine practice, the LRT for this comparison is evaluated as a χ^2 with 3 d.f. (9 parameters in general model – 6 parameters in nested model). However, the null hypothesis here (the mixed codominant model) has three constrained parameters of which two are constrained at their boundaries (since, $0 < \tau < 1$). Hence, the distribution of the LRT is better represented as a 1/4:1/2:1/4 mixture of χ^2 s with 1, 2, and 3

d.f., respectively, which would result in a lower critical value (6.25 vs. 7.81 for a χ^2 with 3 d.f.) corresponding to rejecting the null hypothesis at the 5% significance level. Alternatively or where appropriate mixtures are not known, the LRT test could be evaluated at the upper and lower ranges of the number of estimated parameters with the smaller d.f. representing a more liberal test and the larger d.f. representing a more conservative test [28]. In the interest of comprehensiveness, mixtures of chi-squares and ranges of p values are both utilized for model comparison. Finally, the Akaike Information Criteria (AIC) [29] was used to determine the most parsimonious model (i.e. the model with the smallest value of AIC).

Data Simulation

Based on the results from measured tIgE data, 200 replicates of a quantitative trait were simulated in PAP. In the simulation step of this analysis, all the inbreeding was taken into account by utilizing the full 13-generational pedigree structure to simulate the quantitative trait. In the segregation analysis phase, however, only phenotype data on the original 445 members phenotyped were retained, and analysis was performed using the same restricted pedigree structures utilized in the segregation analyses on observed tIgE. This was done to evaluate the reliability of the segregation analysis on subsets of the full genealogy, i.e. to evaluate the failure to utilize the true pedigree structure. The general model, the model selected based on the observed data (and smaller models as necessary) were fit to each of the 200 replicates. Between 3 (for the smaller models with fewer parameters) and 5 (for the larger models with more parameters) starting point estimates were used for each replicated data set. While one set of initial estimates was selected to be close to the generating values, the others were set to have initial estimates far from the generating values.

Results

Preliminary Data Analysis

Between the years 1995 and 2001 (when subject recruitment occurred), there were 664 individuals on Tangier Island. Total serum IgE was measured on 453 (68.2%) individuals. A total of 139 (21%) subjects refused to par-

ticipate in this study, contact was never established for only 27 subjects, and 7 were potential recruits (i.e. they expressed willingness but a convenient recruitment time was never found). The remaining subjects were ineligible (too young or too old to knowingly consent). A total of 445 subjects were used in the final analysis (eight subjects were dropped as they were not residents). To evaluate the validity of a single time-point measure of tIgE, for 102 subjects blood draws were performed at two different times on average 3 years apart. Following adjustment for age at bleed, the correlation across these repeat samples was excellent (0.99) and the coefficient of variation was very low (average CV = 0.7%, range = 0.1–1.4%). This indicates that beyond the linear effects of age, there is little variation in tIgE levels in these subjects, and for these samples, results from the later of the two blood draw was used for analysis.

There were more female than male participants: 260 (58%) vs. 185 (42%), respectively. The average age at recruitment was 45.6 years (ranging from 6 to 97 years), and did not differ between males and females (44.7 vs. 46.2 years, respectively). Mean tIgE was 164.57 ng/ml with a standard deviation of 357.21 ng/ml. These tIgE levels were highly skewed to the right (skewness coefficient of +5.10 and kurtosis of +33.96). A logarithm (base 10) of tIgE ($\log[tIgE]$) achieved a Gaussian distribution, and no observations are more than 3 standard deviations away from the sample mean.

The geometric mean of tIgE was 56.23 ng/ml, and was higher in males compared to females (81.3 vs. 42.7 ng/ml, $p < 0.001$). Five percent of the total phenotypic variation in $\log[tIgE]$ could be explained by gender, and 3% was explained by age ($p < 0.001$). No statistically significant variation in tIgE levels was observed by month or season of bleed. As expected, mean $\log[tIgE]$ was higher in ever smokers compared to never smokers (geometric mean 67.81 vs. 43.65 ng/ml, respectively, $p = 0.011$). However, the observed relationship between smoking and tIgE in this population could be fully explained by confounding with gender (most smokers were male), and removing gender effects essentially removed all effects of smoking. Based on these results, only effects of age and gender were removed from $\log[tIgE]$ using multiple regression models prior to the segregation analysis to avoid estimating two additional parameters within the modeling process. Adjusted values were centered to males at age 20.

Pedigree Feasibility

Table 1 summarizes the evaluation of feasibility of a segregation analysis. It is apparent that pedigree struc-

tures 1, 2 and 4 were not conducive to a conventional approach, with each maximization taking at least 4 months. The only feasible options that remained were pedigree structures 3 and 5: the inclusion of all 445 subjects with their parents, and nuclear families with duplication of individuals. Based on these results, these two pedigree structures were selected for the final analysis: (1) a *reduced* pedigree structure, including 57 extended families with no duplication of individuals, and (2) the *nuclear* families with the duplication of individuals as needed. Both these pedigree sub-structures ignore all inbreeding.

For the reduced pedigrees, the full 13-generational pedigree was broken down into 57 pedigrees without duplication of any individuals. Since the 445 phenotyped individuals belong to multiple generations (generation 5 to 13) in the original 13-generational pedigree, the retention of these and their parents resulted in 57 pedigrees, the largest of which was 7 generations deep with 443 individuals and 161 sibships, while the smallest pedigree had 3 individuals. The average pedigree size was 13.1 individuals and average depth was 2.5 generations. All 57 pedigrees had at least one phenotyped individual. In the nuclear pedigree structure, the full 13-generational pedigree was broken down into 922 nuclear pedigrees with the duplication of 1,476 individuals. The largest nuclear family had 16 individuals (14 offspring) and smallest had 3 (1 offspring). The average pedigree size was 5.4 individuals. Of these nuclear families only 330 had at least one member with measured IgE. With the duplication of individuals, the total number of individuals with measured IgE was 742 (297 individuals were duplicated). We present here in detail the segregation analysis results based on the reduced pedigree, and only in summary those based on the nuclear pedigrees for three reasons: (1) this reduced pedigree structure is free from the high duplication of phenotyped subjects that otherwise occurred in nuclear families; (2) the reduced pedigree structure retains maximal genealogical information permissible by current computer capacities, and (3) because the results based on the nuclear pedigrees were very similar.

Familial Correlations

Spousal correlation in $\log[tIgE]$ was very close to zero ($r^2 = -0.017$, $n = 106$). Parent-offspring correlation was considerably higher at 0.150 ($p = 0.004$, $n = 377$); however, the correlation between mother and child was higher than that between father and child (0.183, $p = 0.01$, $n = 239$ vs. 0.088, $p = 0.31$, $n = 138$). Correlation was similar to parent-offspring for half-sibs (0.143, $p > 0.5$, $n = 5$). Male-male and female-female sib pairs had correlations (0.160,

Table 2A. Segregation analysis of age- and gender-adjusted log[IgE] using reduced pedigree structure

Model	Parameter (SE)										# par	−2ln(L)	AIC	χ^2 mixtures p value ¹ range p value ²
	q _A	μ _{AA}	μ _{AB}	μ _{BB}	σ	h ²	τ _{AA}	τ _{AB}	τ _{BB}					
A. Sporadic	[100%]	2.029 (0.029)	= μ _{AA}	= μ _{AA}	0.609 (0.020)	[0]	[1.0]	[0.5]	[0.0]	2	820.794	824.794	12.881 0.047 (0.025, 0.075)	
B. Polygenic	[100%]	2.024 (0.036)	= μ _{AA}	= μ _{AA}	0.612 (0.022)	0.261 (0.127)	[1.0]	[0.5]	[0.0]	3	815.071	821.071	7.158 0.258 (0.209, 0.306)	
C. Mixed dominant	0.253 (0.068)	2.418 (0.088)	= μ _{AA}	1.695 (0.084)	0.494 (0.031)	0.000*	[1.0]	[0.5]	[0.0]	5	810.585	820.585	2.672 0.442 (0.263, 0.614)	
D. Mixed codominant	0.769 (0.068)	1.750 (0.084)	2.433 (0.141)	2.466 (0.622)	0.499 (0.033)	0.000*	[1.0]	[0.5]	[0.0]	6	810.562	822.562	2.649 0.271 (0.116, 0.481)	
E. Dominant floating τ _{AB}	0.270 (0.081)	2.431 (0.084)	= μ _{AA}	1.719 (0.080)	0.495 (0.031)	0.000*	[1.0]	0.428 (0.080)	[0.0]	6	809.892	821.892	1.979 0.370 (0.159, 0.577)	
F. Three distribution, non-genetic	0.639 (0.437)	2.097 (0.363)	1.781 (0.306)	2.654 (0.607)	0.544 (0.082)	0.345 (0.197)	q _A	q _A	q _A	6	813.853	825.853	5.94 0.115 –	
G. General	0.317	2.165	2.474	1.684	0.475	0.000	0.289	0.534	0.008	9	807.913	825.913		

par = Number of estimated parameters; * parameter arbitrarily fixed to allow standard error calculation.
¹ p value calculated using mixtures of χ^2 .
² p values for smallest and largest degrees of freedom.

Table 2B. Segregation analysis of age- and gender-adjusted log[IgE] using reduced pedigree structure for two-distribution models

Model	Parameter (SE)										# par	-2ln(L)	AIC	χ^2 mixtures p value ¹ range p value ²
	q _A	μ _{AA}	μ _{AB}	μ _{BB}	σ	h ²	τ _{AA}	τ _{AB}	τ _{BB}					
C. Mixed dominant	0.253 (0.068)	2.418 (0.088)	= μ _{AA}	1.695 (0.084)	0.494 (0.031)	0.000*	[1.0]	[0.5]	[0.0]	5	810.585	820.585	1.717 0.418 (0.190,0.633)	
F2. Two-distribution, non-genetic	0.104 (0.168)	2.545 (0.316)	= μ _{AA}	1.896 (0.178)	0.557 (0.061)	0.325 (0.171)	q _A	q _A	q _A	5	813.897	823.897	5.029 0.170 –	
G2. Two-distribution, general	0.306	1.703	= μ _{AA}	2.445	0.482	0.000	0.313	0.628	0.070	8	808.868	824.868		

par = number of estimated parameters; * parameter arbitrarily fixed to allow standard error calculation.
¹ p value calculated using mixtures of χ^2 .
² p values for smallest and largest degrees of freedom.

$n = 47$ and 0.120 , $n = 80$, respectively) similar to the parent-offspring correlation. Male-female sib pairs showed virtually no correlation (-0.03 , $n = 121$), but this was largely due to five highly discordant pairs.

Analysis Using The Reduced Pedigree

Tables 2A and B summarize results of the segregation analysis of age and gender adjusted $\log[tIgE]$ using the reduced pedigrees. In models where the polygenic heritability settled on the lower boundary (i.e. $\hat{h}^2 = 0$), the maximum likelihood was obtained allowing heritability to vary, but standard errors for other estimated parameters were obtained by fixing residual heritability to a small value (0.001).

Comparing models A through F to the general model G (table 2A) showed that only the sporadic model could be rejected. The substantial familial correlation in $\log[tIgE]$ is reflected in a total additive polygenic heritability of $26 \pm 13\%$ (model B). The mixed dominant model C was not significantly different from the general model ($p = 0.442$). A mixed recessive model fit (data not shown) was the mirror image of the mixed dominant model. Major gene models absorbed most of the familial correlations, with residual additive heritability tending to zero. The extension to a mixed model (D) with three distributions showed no significant improvement over two-distribution models. The LRT comparing models C and D yielded a p -value of 0.879 , and the estimated means for two genotypes AB and BB were almost identical in model D. While the three distribution major-environmental factor with polygenic effects model F also gave a reasonable fit to these data, the AIC favored the two-distribution mixed Mendelian model C as most parsimonious. Additionally, the putative environmental factor in model F could not explain the familial correlations seen in the data. Finally, a dominant model was fit (model E) holding τ_{AA} and τ_{BB} to their Mendelian expectations of 1.0 and 0.5 , respectively, while allowing τ_{AB} to vary. Here, τ_{AB} was estimated to be very close to its Mendelian expectation at 0.428 ± 0.080 .

Two additional models were fit corresponding to models F and G with only two distributions (table 2B). This was done to confirm that a two-distribution major environmental factor with polygenic effects model was not better than a two-distribution Mendelian model. Comparison of these models also favored the two-distribution Mendelian mixed model as the most parsimonious model for the distribution of $tIgE$ in these 57 families.

Briefly, analysis of the 922 nuclear families was similar to that of the 57 reduced pedigrees (data not shown). The

observed familial correlations in adjusted $\log[tIgE]$ were reflected in the polygenic models ($29 \pm 10\%$ in the nuclear families). The distinction between the single-distribution polygenic, and two-distribution mixed model was somewhat clearer with the nuclear pedigrees (AICs were 3 units apart). For both pedigree structures, the most parsimonious model (as indicated by the AIC) was the mixed Mendelian model with two distributions, where the major gene accounted for almost all the genetic variance (residual heritability tended to 0 for the nuclear families also). Here too, larger and non-Mendelian models also could not be rejected.

Reliability of the Model Predicted

The reduced pedigree structure indicated that the most parsimonious model for $\log[tIgE]$ in these data was a two-distribution Mendelian model in which the major gene accounted for most of the genetic variance. Simulations under this two-distribution major-gene-only model were performed to evaluate the reliability of estimated parameters.

Parameter estimates from the segregation analysis on the real data were used in these simulations; specifically – allele frequency of A was set at 0.253 , mean for genotypes AA and AB were set at 2.418 , mean for genotype BB at 1.695 , common standard deviation of 0.494 and there was no residual heritability. Two hundred replicates were generated, and for each replicate, four models were evaluated: (1) the polygenic model B; (2) the two-distribution mixed model C; (3) the general model G, and (4) a general model with τ_{AA} and τ_{BB} fixed at 1 and 0 , respectively. Comparing this second general model (with two fixed transmission probabilities) is fairly routine practice, because the general model tends to maximize with transmission probabilities similar to Mendelian expectations (i.e. $\tau_{AA} = 1$ and $\tau_{BB} = 0$).

Model comparison was performed only on those replicates that were free from major maximization issues (i.e. likelihood of largest model $G >$ likelihood of model $C >$ likelihood of smallest model B). Following standard protocol, a two-step procedure was used to compare models: (1) models B and C were compared to model G using the LRT, and (2) in those replicates where both models B and C were not rejected, the AIC was used to determine parsimony. This testing sequence is illustrated in figure 1.

After excluding replicates with maximization problems, there were 170 successful replicates (table 3). Using the LRT statistic alone, the absolute ability to detect a dominant major gene model (i.e. to find it was not different from the general model) was high (89%). However, in

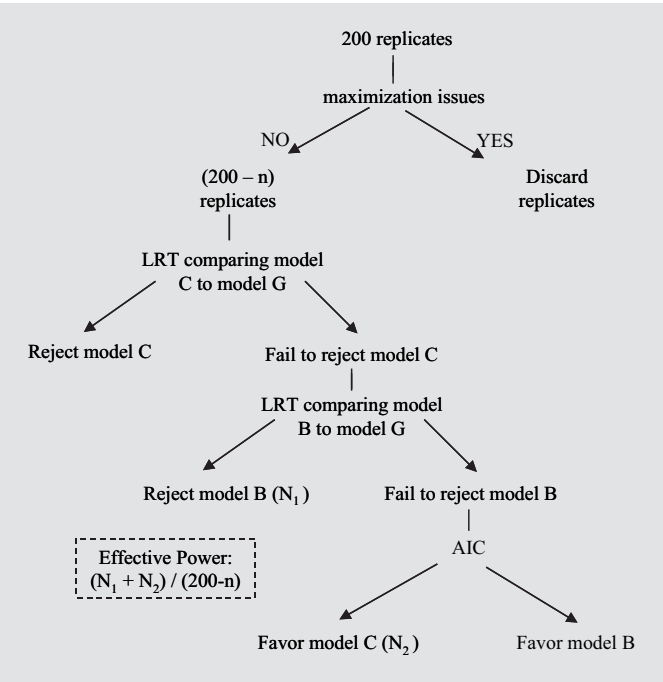


Fig. 1. Testing sequence to evaluate reliability and power of the two-distribution Mendelian model. Model G = General unconstrained model; C = generating two-distribution Mendelian model; B = polygenic model.

the majority of replicates it was not possible to reject the simpler polygenic model based on the LRT statistic alone. The two competing models were further compared using AICs to determine parsimony for these replicates. Of the 114 replicates in which both the two-distribution and single distribution models gave a good fit to the data, 55 replicates had lower AICs for the two-distribution model in the reduced pedigrees, resulting in an overall power of 54% (i.e. 92 out of 170 replicates successfully identified the major gene).

Parameter estimates appeared significantly different from the generating values as indicated by the t-tests ($p < 0.001$) comparing the average MLE to the generating value (table 3). Allele frequency appeared to be overestimated, the ‘dominant’ mean underestimated, and the mean of the recessive homozygote overestimated. This, however, is due to the estimated mean of heterozygotes (i.e., dominant vs. recessive for ‘high’ trait values). These data were simulated under a model where heterozygotes had the higher mean (i.e. a dominant model), however in 46 of the 151 replicates where the two-distribution model was found to have good fit, heterozygotes were esti-

Table 3. Summary from 200 replicates for a dominant two-distribution Mendelian model with no residual heritability

Summary statistic	Reduced pedigree
Number replicates in final stage	170
Detecting two-distribution model	151
LRT fails to reject polygenic model	114
AIC favors two-distribution Mendelian model	55
Selecting true generating model	92
Average parameter estimates ($\pm$ SD)	
q_A	0.31 ($\pm$ 0.16)
μ_{AA}, μ_{AB}	2.19 ($\pm$ 0.33)
μ_{aa}	1.93 ($\pm$ 0.48)
h^2	0.05 ($\pm$ 0.10)
Heritability from polygenic model ($\pm$ SD)	0.29 ($\pm$ 0.10)
Models identified as recessive	30.46%

mated to have the lower mean (i.e. the complementary recessive model was selected). Predicted proportions of individuals in the higher and lower distributions remained constant, since the estimated allele frequency changed accordingly. This is consistent with the direction of bias in the average MLEs described above. In fact, restricting this evaluation to the 105 replicates in which the heterozygote had the higher mean, the MLEs were found to be not different from their generating values. We should highlight also that this reversal of the heterozygote mean in the simulations, may be a result of the limited number of starting points used here. While 3–5 sets of initial estimates are generally inadequate to search a likelihood surface, for the simulations attempted here only this limited number was feasible. Finally, the estimated total additive heritability in the polygenic models was 0.29 ($\pm$ 0.10).

Discussion

Both asthma and tIgE levels are complex multifactorial phenotypes. It is the general consensus that neither trait is controlled by a simple genetic mechanism, but rather a complex interplay of multiple genes, exposure to environmental components (most relevant being allergens), the interactions between these two, along with the time of exposure leading first to sensitization in early life and later acting as triggers. One advantage of studying a

population that is geographically isolated and restricted to a small area, such as Tangier Island, is that the 'macro' environment may be relatively uniform for allergen exposure, lifestyle and diet. Furthermore, this population is sufficiently inbred to suggest that the gene pool of the current residents is somewhat more homogeneous. Finally, perhaps the strongest advantage with regard to these data is the population-based recruitment scheme, where all residents were invited to participate in the study without regard to disease status. Of the 664 Tangier Island residents between the years 1995 and 2002, only 27 could not be contacted. Total serum IgE was measured on a total of 453 individuals (68.2% of the total population). This ascertainment scheme removes the biases that arise due to selection of families enriched for the trait or a highly correlated trait (asthma or allergic rhinitis in the case of IgE), a particular problem for segregation analysis. We speculate that the aforementioned strengths would result in greater homogeneity, and hence, an ample dataset (445 phenotyped individuals in 57 apparently independent pedigrees) for our segregation analysis.

Evidence to support genetic control of total serum IgE in this Tangier Island population is compelling based both on the observed familial correlations and estimated heritability in these analyses. In these data the correlation between mother-offspring pairs was higher than father-offspring suggesting a parent-of-origin effect. There is a limited amount of literature supporting a maternal pattern of inheritance for atopy and high IgE levels [30, 31]. However, the greater environmental similarity between mother-child compared to father-child should not be ignored, as this could also lead to higher correlations for mother-offspring pairs. This is certainly plausible in the situation of Tangier Island where ethnographic observations indicate a relatively more sedentary lifestyle observed for women, young adults and children in contrast to the more active lifestyle of the watermen.

Given the quantity of genealogical information available on the Tangier Island population, it would have been optimal to perform a conventional segregation analysis using the full 13-generational pedigree. The limiting step for segregation analysis is not in the availability of algorithms to handle pedigrees of large sizes. Pedigree Analysis Package (PAP) [25] which approximates the likelihoods for segregation analysis and Monte Carlo Markov chain (MCMC) methods implemented in MORGAN [32] or LOKI [33] can handle pedigree of large sizes. Our feasibility study demonstrates that running analyses on the full pedigree structure in PAP is clearly not practical given its complexity, and MCMC methods while requiring

fewer sets of initials are not expected to be feasible on pedigrees with much more than 10 loops (personal communication, Dr. Ellen J. Wijsman). Thus, the limiting step for this large pedigree is its complexity, resulting in a run time for a single set of initial estimates which would be in excess of six months for the simple polygenic model. Since multiple sets of initials are then required for many models in a segregation analysis, analysis of the full pedigree was not feasible.

In the context of linkage analysis, ignoring inbreeding typically has some consequences, the direction of which is determined by the particular analytical approach being implemented [34–38]. Furthermore, it has been shown that the choice of pedigree trimming also has an effect on linkage analysis [37, 39, 40]. Various ways to split large pedigrees into smaller analyzable sub-pedigrees have been attempted, encompassing systematic rules based solely on pedigree structure [37, 40], ad hoc rules to enrich for selected phenotypes [39], and methods that maximize information for linkage [41]. In the absence of any algorithm to split this pedigree to maximize information for segregation analysis, we implemented systematic rules based on pedigree structure: removing upper layers of the pedigree starting with the earliest generation, or adding layers to the pedigree starting with the phenotyped individuals. In removing the top five generations of the pedigree, 1,655 individuals of whom only 17 were phenotyped were removed. Removal of the sixth generation, while removing an additional 890 individuals, would have also lost 153 phenotyped subjects, and hence we removed only the upper five generations. In adding generations from the bottom, since the phenotyped individuals belonged to multiple generations in the original pedigree, adding one and two generations above these phenotyped subjects, resulted in 57 and 19 *apparently* independent pedigrees, respectively. Analysis was then performed using maximal genealogical information permissible under current computer limitations.

As in the case of linkage analysis where ignoring inbreeding generally has some impact, the selection of models in segregation analyses may be affected to an unknown extent by ignoring deeper levels of genealogical complexity. Two parameters in a segregation analysis likely to be most biased when inbreeding is ignored are allele frequencies and heritability estimates since these are highly dependent on pedigree structure. In an inbred population, ignoring inbreeding could lead to an overestimation of heritability, and an overestimation of the heterozygote frequency [42], but given the low levels of inbreeding in this population, the level of bias is arguably minimal.

Based on the polygenic models from the segregation analysis, the narrow sense heritability for total serum IgE in this population was 26–29%. To evaluate potential bias, simulations for a single distribution polygenic model were carried out for the Tangier Island pedigree structure (data not shown) similar to the methods described above for the major gene model. Briefly, utilizing the complete genealogy, 200 replicates of a quantitative trait under two different polygenic models (heritability = 26 and 40%) were simulated. Heritability was then computed retaining information for only the original phenotyped subjects and using only the 57 reduced pedigrees. Ignoring inbreeding had little to no impact on estimates for heritability for the two models (0.249 ± 0.086 and 0.389 ± 0.103 , respectively), and these parameter estimates appeared to be unbiased (*p* value comparing MLE to the generating value = 0.082 and 0.172, respectively). With respect to allele frequencies, Hardy-Weinberg equilibrium was always assumed in the analyses presented here, which calls for the allele frequencies to stay constant across generations. However, where inbreeding is present, there is a correlation among uniting gametes that reflects average levels of inbreeding which does not change the allele frequency itself across generations, but reduces the expected number of heterozygotes [43]. When *F* is low, this fractional decrease would be minimal and should have little impact on the segregation analysis performed here, as seen in the simulations under the major gene two-distribution model.

In the segregation analyses performed here, the hypotheses tested consisted of a single general model in which all parameters were estimated, and various restricted models including mixed Mendelian models with two- and three-distributions. While none of the genetic and non-genetic models could be definitely rejected, a two-distribution model was found most parsimonious. The addition of a third genotypic mean in the three-distribution model showed no improvement in fit. Moreover, AIC favored the two-distribution Mendelian model as most parsimonious. This model was also favored over the two-distribution equal-transmission model, indicating that the transmission of the phenotype from parent to offspring was under genetic control and not purely environmental, even though the environmental model could not be rejected. The major gene implicated here absorbed all the familial aggregation in $\log[\text{tIgE}]$, and major-gene only models also supported these results (data not shown).

Results of simulations where the quantitative trait was simulated using full genealogy with subsequent analysis

restricted to subsets used in the real data, were very similar to those obtained with the real data. This reveals a minimal impact of underestimating inbreeding. For the simulated two-distribution model, using only the LRT statistic, a polygenic model was found to be no different from the general model for the majority of replicates, as we observed with the real data. Estimates of polygenic heritability averaged 29%, similar to estimated polygenic heritability in the real data and appeared to be unbiased. For a majority of these replicates, both polygenic and two-distribution models could not be rejected using the LRT and incorporating the AICs into the testing strategy to identify the best fitting and most parsimonious model in these replicates, clearly favored the generating two-distribution major-gene-only model, and in general residual heritability tended to zero. These observations are consistent with the expectation of a minimal impact of underestimating inbreeding on the estimates of heritability and allele frequencies in the segregation analyses as discussed above.

Thus the most parsimonious model from the segregation analysis suggests total serum IgE in the Tangier Island population appears to have essentially two underlying genotypic distributions, with a slight majority of founders in the 'lower' distribution. The mean of the higher distribution was 2.42 $\log(\text{ng/ml})$ or 261.82 ng/ml for age- and gender-adjusted IgE, and roughly 44% of founders were predicted to fall in this distribution. The estimated mean of the lower distribution was 1.70 $\log(\text{ng/ml})$ or 49.55 ng/ml for age- and gender-adjusted IgE, and a predicted 56% of the founders fall in this distribution.

Results of the segregation analysis of tIgE in this Tangier Island population are similar to published studies. Ten of the past twelve studies on inheritance of total serum IgE favored some major gene model. Of these, six found that a two-distribution model was best [10–13, 20, 21], two found the best fitting model was a three-distribution model [15, 17], and two determined a mixed model [18, 19] (both with two distributions from the major gene component) to be best. While the two-distribution models are most consistently supported, their predicted distributions varied considerably. In many of the studies, the vast majority of individuals fall into one distribution [12, 13, 18, 19]. In contrast, results from the Tangier Island population predict that individuals are almost equally divided between the two distributions. The French study [21] showed similar, almost equal division, between the high and low predicted distributions. What is unique about the most parsimonious model from the Tangier Island data is that it implicates only a major gene effect. In fitting

mixed models, the putative major gene accounted for virtually all the genetic variation. This is different from past studies where residual familial correlations were clearly present. For those studies that relied on the regressive classes of models, statistically significant residual correlations between relative pairs were observed [12, 13, 17, 21], and those that relied on mixed models found residual polygenic heritability to be present [18, 19]. It should be noted that while the two-distribution Mendelian model was most parsimonious, non-genetic two distribution models could not be rejected. Hence, while our data do not conclusively establish that a major gene alone is responsible for tIgE control in the Tangier Island population, it suggests that tIgE is under some genetic control perhaps as part of a complex physiologic network of interacting causal agents.

In conclusion, it appears that in Tangier Island, a two-distribution major gene model is the most parsimonious

explanation for the distribution of total serum IgE, and a major gene may account for most of the genetic variation in this quantitative trait. The complexity of the pedigree made it unfeasible to analyze the complete genealogy, but use of restricted sub-pedigrees in a segregation analysis appeared to have little impact on the best fitting model.

Acknowledgements

We thank the families from Tangier Island, Virginia, for their wholehearted cooperation and participation in this study, especially Inez Pruitt, Jean Crockett and Cindy Parks for their local support. We thank Mark Liu, Eva Ehrlich, Renate Nickel, Lun Xiu, Kathryn Held, Sharon Patterson and April Zambelli-Weiner for their support in study design, subject recruitment and sample processing. This study was funded by the Center for Indoor Air research with approval by the Bayview Institutional Review Board (AAC98-01-01-01) at Johns Hopkins University.

References

- ▶ 1 Woolcock AJ, Peat JK: Evidence for the increase in asthma worldwide. *Ciba Found Symp* 1997;206:122–34.
- ▶ 2 Jarvis D, Burney P: ABC of allergies. The epidemiology of allergic disease. *BMJ* 1998;316:607–610.
- ▶ 3 ECRHS: Variations in the prevalence of respiratory symptoms, self-reported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey (ECRHS). *Eur Respir J* 1996;9:687–695.
- ▶ 4 Burrows B, Martinez FD, Haonen M, Barbee RA, Cline MG: Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N Engl J Med* 1989;320:271–277.
- ▶ 5 Sporik R, Ingram JM, Price W, Sussman JH, Honsinger RW, Platts-Mills TA: Association of asthma with serum IgE and skin test reactivity to allergens among children living at high altitude. Tickling the dragon's breath. *Am J Respir Crit Care Med* 1995;151:1388–1392.
- ▶ 6 Wüthrich B, Baumann E, Fries RA, Schnyder UW: Total and specific IgE (RAST) in atopic twins. *Clin Allergy* 1981;11:147–154.
- ▶ 7 Hanson B, McGue M, Roitman-Johnson B, Segal NL, Bouchard TJ Jr, Blumenthal MN: Atopic disease and immunoglobulin E in twins reared apart and together. *Am J Hum Genet* 1991;48:873–879.
- ▶ 8 Strachan DP, Wong HJ, Spector TD: Concordance and interrelationship of atopic diseases and markers of allergic sensitization among adult female twins. *J Allergy Clin Immunol* 2001;108:901–907.
- ▶ 9 Jacobsen HP, Herskind AM, Nielsen BW, Husby S: IgE in unselected like-sexed monozygotic and dizygotic twins at birth and at 6 to 9 years of age: High but dissimilar genetic influence on IgE levels. *J Allergy Clin Immunol* 2001;107:659–663.
- ▶ 10 Marsh DG, Bias WB, Ishizaka K: Genetic control of basal serum immunoglobulin E level and its effect on specific reaginic sensitivity. *Proc Natl Acad Sci USA* 1974;71:3588–3592.
- ▶ 11 Gerrard JW, Rao DC, Morton NE: A genetic study of immunoglobulin E. *Am J Hum Genet* 1978;30:46–58.
- ▶ 12 Meyers DA, Beaty TH, Colyer CR, Marsh DG: Genetics of total serum IgE levels: A regressive model approach to segregation analysis. *Genet Epidemiol* 1991;8:351–359.
- ▶ 13 Dizier MH, Hill M, James A, Faux J, Ryan G, le Souef P, Lathrop M, Musk AW, Demenais F, Cookson W: Detection of a recessive major gene for high IgE levels acting independently of specific response to allergens. *Genet Epidemiol* 1995;12:93–105.
- ▶ 14 Hasstedt SJ, Meyers DA, Marsh DG: Inheritance of immunoglobulin E: genetic model fitting. *Am J Med Genet* 1983;14:61–66.
- ▶ 15 Meyers DA, Bias WB, Marsh DG: A genetic study of total IgE levels in the Amish. *Hum Hered* 1982;32:15–23.
- ▶ 16 Borecki IB, Rao DC, Lalouel JM, McGue M, Gerrard JW: Demonstration of a common major gene with pleiotropic effects on immunoglobulin E levels and allergy. *Genet Epidemiol* 1985;2:327–338.
- ▶ 17 Martinez FD, Holberg CJ, Halonen M, Morgan WJ, Wright AL, Taussig LM: Evidence for Mendelian inheritance of serum IgE levels in Hispanic and non-Hispanic white families. *Am J Hum Genet* 1994;55:555–565.
- ▶ 18 Meyers DA, Beaty TH, Freidhoff LR, Marsh DG: Inheritance of total serum IgE (basal levels) in man. *Am J Hum Genet* 1987;41:51–62.
- ▶ 19 Xu J, Postma DS, Howard TD, Koppelman GH, Zheng SL, Stine OC, Bleecker ER, Meyers DA: Major Genes Regulating Total Serum Immunoglobulin E Levels in Families with Asthma. *Am J Hum Genet* 2000;67:1163–1173.
- ▶ 20 Blumenthal MN, Namboodiri K, Mendell N, Gleich G, Elston RC, Yunis E: Genetic transmission of serum IgE Levels. *Am J Med Genet* 1981;10:219–228.
- ▶ 21 Sampogna F, Demenais F, Hochez J, Oryszczyn MP, Maccario J, Kauffmann F, Feingold J, Dizier MH: Segregation analysis of IgE levels in 335 French families (EGEA) using different strategies to correct for the ascertainment through a correlated trait (asthma). *Genet Epidemiol* 2000;18:128–142.
- ▶ 22 Parks T, Felix K, Rice T, Subbarao PV, Marimuthu KM, Rao DC: A genetic study of immunoglobulin E and atopic disease based on families ascertained through asthmatic children. *Hum Hered* 1990;40:69–76.
- ▶ 23 Mathias RA, Bickel CA, Beaty TH, Petersen GM, Hetmanski JB, Liang KY, Barnes KC: A study of contemporary levels and temporal trends in inbreeding in the Tangier Island, Virginia, population using pedigree data and isonymy. *Am J Phys Anthropol* 2000;112:29–38.

- ▶ 24 Maclean CJ, Morton NE, Elston RC, Yee S: Skewness in commingled distributions. *Biometrics* 1976;32:695–699.
- 25 Hasstedt SJ, Cartwright PE: PAP: Pedigree Analysis Package. No. 13. Salt Lake City, University of Utah, 1981.
- 26 Edwards A: Likelihood: An account of the statistical concept of likelihood and its application to scientific inference, 2nd ed. London, Cambridge University Press, 1978.
- ▶ 27 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.
- ▶ 28 Elston RC: Segregation analysis. *Adv Hum Genet* 1981;11:63–120.
- 29 Akaike H: A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;AC-19:716–723.
- ▶ 30 Cookson WO, Young RP, Sandford AJ, Mofatt MF, Shirakawa T, Sharp PA, Faux JA, Julier C, Nakumuura Y, Nakumura Y: Maternal inheritance of atopic IgE responsiveness on chromosome 11q. *Lancet* 1992;340:381–384.
- ▶ 31 Holberg CJ, Morgan WJ, Wright AL, Martinez FD: Differences in familial segregation of FEV1 between asthmatic and nonasthmatic families. Role of a maternal component. *Am J Respir Crit Care Med* 1998;158:162–169.
- ▶ 32 Snow GL, Wijsman EM: Pedigree analysis package (PAP) vs. MORGAN: model selection and hypothesis testing on a large pedigree. *Genet Epidemiol* 1998;15:355–369.
- ▶ 33 Heath SC: Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997;61:748–760.
- ▶ 34 Genin E, Clerget-Darpoux F: Consanguinity and the sib-pair method: an approach using identity by descent between and within individuals. *Am J Hum Genet* 1996;59:1149–1162.
- ▶ 35 Genin E, Clerget-Darpoux F: Association studies in consanguineous populations. *Am J Hum Genet* 1996;58:861–866.
- ▶ 36 Genin E, Todorov AA, Clerget-Darpoux F: Properties of the transmission-disequilibrium test in the presence of inbreeding. *Genet Epidemiol* 2002;22:116–127.
- 37 Dyer TD, Blangero J, Williams JT, Goring HHH, Mahaney MC: The effect of pedigree complexity on quantitative trait linkage analysis. *Genet Epidemiol* 2001;21(suppl 1):236–243.
- ▶ 38 Leutenegger AL, Genin E, Thompson EA, Clerget-Darpoux F: Impact of parental relationships in maximum lod score affected sib-pair method. *Genet Epidemiol* 2002;23:413–425.
- 39 Chapman NH, Leutenegger AL, Badzioch MD, Mogdan M, E.M. C, Daw EW, Gagnon F, Li N, Maia JM, Wijsman EM, Thompson EA: The importance of connections: Joining components of the Hutterite Pedigree. *Genet Epidemiol* 2001;21(suppl 1):230–235.
- 40 Pankratz VS, Iturria SJ: A pedigree partitioning approach to quantitative trait loci mapping of IgE serum level in the GAW12 Hutterite data. *Genet Epidemiol* 2001;21(suppl 1):S258–S263.
- ▶ 41 Falchi M, Forabosco P, Mocchi E, Borlino CC, Picciau A, Virdis E, Persico I, Parracciani D, Angius A, Pirastu M: A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia. *Am J Hum Genet* 2004;75:1015–1031.
- 42 Falconer DS, Mackay TFC: Introduction to quantitative genetics. 4th ed. Essex, England: Addison Wesley Longman Ltd., 1996.
- 43 Li CC: Population Genetics. 6th ed. Chicago, The University of Chicago Press, 1968.