A Genetical Model for Vitiligo

Partha P. Majumder,*† S. K. Das,† and C. C. Li*

*Human Genetics Division, Department of Biostatistics, University of Pittsburgh, Pittsburgh; and †Indian Statistical Institute, Calcutta

Summary

A genetical model is found to provide a good fit to family data on vitiligo. The model postulates that recessive alleles at a set of four unlinked diallelic loci are involved in the causation of the disorder. Under this multiple recessive homozygosis model, for normal × affected families ascertained through the affected parent, the expected segregation probability is .063; the estimated value is .053, which is not significantly different from the expected value. For normal × normal families ascertained through an affected offspring, the expected segregation probability is .037; the estimated value is .04.

Introduction

Vitiligo is an idiopathic hypomelanositic dermatological disorder that is characterized by pale, milk-white macules that tend to become progressive over time (Mosher et al. 1979). Although universal in occurrence, its prevalence varies considerably among different countries and ethnic groups. Primarily on the basis of clinical records of hospitals and skin clinics, the prevalence is estimated to be ~1% in Egypt, ~0.39% in Switzerland, ~0.14% in Russia, ~0.24% in London, ~1.64% in Japan, and ~1% in the United States (Lerner 1959; El-Mofty 1968; Indian Chemical Society 1984). On the basis of population surveys, in the Isle of Bornholm in Denmark the prevalence is estimated to be ~0.38% (Howitz et al. 1977), and in western India the prevalence varies from near absence to ~3.6% among different communities (Mehta et al. 1973).

Positive family history and familial aggregation of vitiligo have been noted for a long time (Cockayne 1933; Merelender and Rywlin 1940; Behl 1955; Levai 1958; Lerner 1959; Mehta et al. 1973; Goudie et al. 1980; Hafez et al. 1983). Both members of two

monozygotic twin pairs have also been found to be affected (Mohr 1951; Siemens 1953). The mode of inheritance of the disorder is still debated. Some believe that vitiligo is due to an incompletely penetrant autosomal dominant gene (Cockayne 1933; Lerner 1959), while some others claim that the disorder gene is autosomal recessive or that the disorder is polygenic/multifactorial in nature (Mehta et al. 1973; Carnevale et al. 1980). Systematic family studies of this disorder are, however, sadly lacking. Recently, Hafez et al. (1983) have studied families of 150 vitiligo patients. Their data did not support either a single-locus autosomal dominant or an autosomal recessive model, but a multifactorial model provided an adequate fit. The heritability was estimated to be 72.4%. Hafez et al. (1983), however, did not account for variable age at onset of the disorder, and therefore their conclusions are to be accepted as tentative.

With a view to understanding the epidemiology and genetics of vitiligo, we had recently undertaken a study in Calcutta, India. In this study, epidemiological data were gathered from 15,685 individuals (7,555 males and 8,130 females). Data on 298 families, each ascertained through a single affected individual, were also gathered. Details of the procedures of data collection are available in Das et al. (1985a, 1985b). Analysis of the epidemiological data revealed the following facts that are pertinent to the present study (for details, see Das et al. 1985a): (1) the prevalence of vitiligo is 0.459%; (2) there are no statistically significant sex or age differences in preva-

lence rates; (3) there is approximately a 4.5-fold increase in prevalence among close biological relatives of affected individuals; and, (4) the age at onset of the disorder is variable—the mean and modal ages at onset are approximately 22 years and approximately 15 years, respectively. Analysis of these family data led to the following conclusions (for details, see Das et al. 1985b): (1) there is a significant familial aggregation; (2) the disorder does not appear to be inherited in a simple dominant or recessive fashion; and, (3) the heritability of liability to the disorder is 46% ± 5% when a polygenic liability threshold model is assumed.

As mentioned in Das et al. (1985b), one of the major problems in the fitting of concrete genetic models to our family data set on vitiligo has been the lack of multiplex families. Only ~10% of the families are multiplex, and, even among multiplex families, most families have only one affected relative of the proband. The parents of affected children are, in a majority of cases, both unaffected, but there is not even a single case of consanguineous mating. Recurrence risks among offspring are much lower than the 25% or 50% values expected, respectively, under single-gene recessive or dominant models (Das et al. 1985b). The disorder, therefore, does not seem to be due to a single gene, either dominant or recessive.

Recently, Li (1987) has suggested a concrete genetic model for a trait for which the vast majority of families will have only one affected member. The purpose of the present study is to investigate whether the model proposed by Li (1987) provides an adequate fit to the vitiligo family data set.

The Family Data Set

Each family in this data set is nuclear, and each was ascertained through a single affected proband. The probands were affected individuals who came for medical treatment/consultation to the Dermatology Department of the Calcutta National Medical College Hospital and the Calcutta Skin Institute. Clinical examinations of family members were performed by a professional dermatologist. In all, data on 298 families were collected. However, because of incomplete information, primarily on ages of individuals, data on 24 families were discarded, and the data on the remaining 274 families have been analyzed in the present study. Of the 274 families under consideration, 86 families were ascertained through an affected parent and the remaining 188 families were

ascertained through an affected offspring. The parental mating type for every family ascertained through an affected parent was normal × affected, while that for every family ascertained through an affected offspring was normal × normal. Eight families (9.3%) ascertained through an affected parent had an affected offspring; the remaining 78 families (90.7%) had only normal offspring. The average sibship size of families ascertained through an affected parent is 2.92 ± 0.26. Of the 188 families ascertained through an affected offspring, 13 families (7%) were multiplex; in the remaining 175 families (93%), the proband had no affected sib. The average sibship size in these 188 families is 4.47 ± 0.15. In the 13 multiplex families, the total number of affected sibs of the 13 probands was 16.

As mentioned earlier, the prevalence of the disorder in the general population from which the families were sampled is .00459. The disorder has a variable age at onset. The distribution of age at onset was estimated from data available on the probands; this distribution is presented in table 1.

The Multiple Recessive Homozygosis Model

This model, proposed by Li (1987), assumes that the trait is due to multiple recessive homozygosis. Thus, if there are k unlinked loci involved in the determination of the trait and if at each locus there are two alleles—A,a; B,b; C,c; etc.—affected individuals are of genotype aabbcc . . . ; individuals of all other genotypes (which have at least one gene denoted by an uppercase letter) are phenotypically normal. Thus, of the 3^k genotypes, only one genotype leads to an affected phenotype, the remaining $3^k - 1$ genotypes being phenotypically normal. If q_i denotes the fre-

Table 1

Distribution of Age at Onset of Vitiligo among Affected Probands

Age at Onset (Years)	Proportion of Probands	Cumulative Proportion $= p_i$	1 − Cumulative Proportion $= z_i$
≤ 10	.283	.283	.717
11–20	.286	.569	.431
21–30	.185	.754	.246
31–40	.108	.862	.138
41–50	.071	.933	.067
51–60	.050	.983	.017
≥ 61	.017	1.000	.000

quency of the recessive allele (the allele denoted by a lowercase letter) at the ith locus ($i = 1, 2, \ldots, k$), then the frequency of the trait in the general population is

$$Q^2 = \prod_{i=1}^{k} q_i^2 \ .$$

Li (1987) has shown that the probability that a normal $\times$ affected family, ascertained through the affected parent, produces an affected offspring is $S_1 = Q/(1 + Q)$.

To obtain the probability that a normal $\times$ normal family, ascertained through an affected offspring, produces an affected offspring is more difficult. To simplify the algebra, we assume that $q_i = q$, for all $i = 1, 2, \ldots, k$. We note that each of the normal parents in a family ascertained through an affected offspring must be either heterozygous or recessive homozygous at each of the k loci and must be heterozygous at least at one locus. This is because in order to produce an aabbcc . . . offspring, each parent must be capable of transmitting an abc . . . gamete, and the reason why neither parent can be recessive homozygote at all the k loci is that each parent is known to be phenotypically normal. If Hardy-Weinberg equilibrium is assumed, the frequencies of heterozygotes and recessive homozygotes are, for each locus, $2q(1 - q)$ $= H$ (say) and $q^2 = R$ (say), respectively. Hence, for any such family in which the father is heterozygous at i loci and the mother is heterozygous at j loci, the mating frequency M_{ij} is

$$M = \frac{\binom{k}{i}\binom{k}{j} H^{i+j} R^{2k-i-j}}{\sum_{i=1}^{k} \sum_{j=1}^{k} \binom{k}{i}\binom{k}{j} H^{i+j} R^{2k-i-j}}$$

The probability that this family produces an affected offspring, θ_{ij}, is

$$\theta_{ij} = \frac{1}{2^{i+j}} \ .$$

Hence, the overall segregation probability, θ, in all such families is

$$\theta = \sum_{i=1}^{k} \sum_{j=1}^{k} M_{ij}\, \theta_{ij} \ .$$

which is another form of Li's (1987) S_2^*. Obviously, q and θ are both dependent on the number of loci (k) and on the prevalence rate in the general population.

Table 2

Change in Gene Frequency (q) and Segregation Probability (θ) in Families Capable of Producing an Affected Offspring with Increase in the Number of Loci (k) under the Multiple Recessive Homozygosis Model when the Prevalence Rate is .00459

k	Gene Frequency $[q_i = q = (0.00459)^{1/2k}]$	Segregation Probability $(\theta = \sum \sum M_{ij}\theta_{ij})$
1	.0677	.2500
2	.2603	.0993
3	.4077	.0552
4	.5102	.0368
5	.5837	.0274

For the given prevalence rate of vitiligo ($= .00459$), in table 2 are presented the values of q and θ for various values of k. As is seen from this table, for a fixed prevalence rate, q increases with k while θ decreases with increase in k.

Analysis of Data of Normal $\times$ Affected Families, Each Ascertained through an Affected Parent

As mentioned in the previous section, the segregation probability in normal $\times$ affected families each ascertained through an affected parent is

$$S_1 = \frac{Q}{1 + Q},$$

where Q^2 is the prevalence rate of the disorder in the general population. In the present case, $Q^2 = .00459$, and $Q = .06775$; therefore, $S_1 = .06345$.

Estimation of the segregation probability from the family data poses a problem because of variable age at onset of the disorder. Since the onset age is variable, normal offspring may actually be of the affected genotype (aabbcc . . .) but may not have manifested the disorder. However, one can estimate the probability, z_x, that an individual of age x is phenotypically normal given that she or he is of the affected genotype (aabbcc . . .). Unfortunately, because of a limited sample size, estimates of z_x are not available for every age x. What we have been able to estimate are the probabilities, p_i ($i = 1, 2, \ldots, 7$), that an individual belonging to an age group i will be affected given that she or he is of the affected genotype (table 1, col. 3). From these estimates, we have estimated the z_x values for the different age groups, z_i ($i = 1, 2,$

..., 7), values that are presented in the last column of table 1. Noting that the probability that an offspring of a normal × affected mating will be of the affected genotype is S_1, and that the probability that an offspring of age x remaining phenotypically normal given that she or he is of the affected genotype is z_∞, one can write down the likelihood of the phenotypic observations on offspring of all the normal × affected families. Suppose, in all these families, that M denotes the total number of offspring and that R denotes the observed total number of affected offspring. Then the likelihood, $L(S_1)$, of the data is

$$L(S_1) \propto \prod_{m=1}^{R} S_1 \cdot (1 - z_{x,m})$$

$$\times \prod_{m=R+1}^{M} [(1 - S_1) + S_1 \cdot z_{x,m}],$$

where $z_{x,m}$ denotes the probability that the mth offspring of age x is of the normal phenotype given that she or he is of the affected genotype. Solving the equation

$$\frac{d \ln L(S_1)}{d S_1} = \frac{R}{S_1} - \sum_{m=R+1}^{M} \frac{1 - z_{x,m}}{1 - S_1 + S_1 \cdot z_{x,m}} = 0,$$

one obtains

$$S_1 = \frac{R}{\displaystyle\sum_{m=R+1}^{M} \frac{1}{(1 - z_{x,m})^{-1} - S_1}} \ .$$

Since, as mentioned earlier, the estimates of the values of z_x are available only for seven age groups and not for individual ages x, the above equation can be rewritten as

$$S_1 = \frac{R}{\displaystyle\sum_{i=1}^{7} u_i \left[\frac{1}{(1 - z_i)^{-1} - S_1} \right]} \ ,$$

where u_i denotes the number of phenotypically normal offspring in age group i ($i = 1, 2, \ldots, 7$). This equation can be solved for S_1 in an iterative manner, starting with an initial value of S_1, to yield the maximum likelihood estimate of S_1.

When this method is used on the normal × affected family data set of vitiligo (comprising 86 families with 251 offspring of whom eight are affected), the maximum likelihood estimate of S_1 turns

out to be .053 ± .0002. This estimated value does not differ significantly (at the 5% level) from the expected value of .063 ($\hat{Z} = [.053 - .063]/\sqrt{.0002} = -.707$). In other words, the multiple recessive homozygosis model provides an adequate fit to the data of normal × affected families, each family ascertained through an affected parent.

Analysis of Data of Normal × Normal Families, Each Ascertained through an Affected Offspring

For normal × normal families capable of producing an affected offspring, the segregation probabilities are given in table 2 for various values of the number of loci, under the assumption of equal gene frequencies at all loci.

To estimate the segregation probability from the data, we adopt a likelihood approach to take into account the variable age at onset of vitiligo and to correct for the bias of ascertainment (since the families are ascertained through an affected offspring). Following Elandt-Johnson (1971), we first outline below the method of calculating the likelihood of one normal × normal family ascertained through an affected offspring. Let:

θ = segregation probability = Prob (offspring is aabbcc . . .);

π = ascertainment probability = Prob (an offspring is a proband given that she or he is affected);

n_i = total number of offspring in age group i;

r_i = observed number of affected offspring in age group i;

$$n = \sum_{i=1}^{I} n_i = \text{total number of offspring in the}$$

family;

$$r = \sum_{i=1}^{I} r_i = \text{total observed number of affected}$$

offspring in the family;

p_i = Prob (an individual in age group i is affected given that she or he is aabbcc . . .).

Therefore, Prob (an offspring in age group i is affected) = $p_i\theta$. Now, Prob (a family with r affected offspring will have at least one proband) = $\alpha_r = 1 -$

Table 3

ln-Likelihood Values of 188 Normal × Normal Families Each Ascertained through an Affected Offspring, for Different Values of Segregation (θ) and Ascertainment (π) Probabilities

	π				Assuming Single Selection
θ	.01	.05	.10	.30	
.01	−167.58	−167.82	−168.13	−169.48	−167.52
.02	−160.49	−160.66	−160.86	−161.81	−160.45
.03	−158.04	−158.12	−158.23	−158.77	−158.02
.04	−157.50	−157.50	−157.51	−157.64	−157.51
.05	−158.03	−157.95	−157.85	−157.57	−158.05
.06	−159.24	−159.07	−158.87	−158.19	−159.28
.10	−167.87	−167.38	−166.77	−164.46	−167.99

$(1 - \pi)^r$. Hence, the probability that in a family with n_i offspring in the ith age group there will be r_i affected offspring in the ith age group is

$$\prod_{i=1}^{g} \binom{n_i}{r_i} (p_i\theta)^{r_i} (1 - p_i\theta)^{n_i - r_i} = P(n_i, r_i) .$$

The probability that such a family is ascertained is $\alpha_r \cdot P(n_i, r_i) = N$. Now, the probability that a family with n_i offspring in the ith age group ($\sum_{i=1}^{g} n_i = n$) has at least one affected offspring and is ascertained is

$$\sum_{r=1}^{n} \left[\alpha_r \cdot \sum_{\substack{(l_1, l_2, \ldots, l_g) \\ l_i \leq n_i \\ \Sigma l_i = r}} P(n_i, l_i) \right] = D .$$

Hence, the conditional distribution of affected offspring in this family when it has been ascertained through an affected offspring (which is the likelihood of this family) is

$$L(\theta, \pi) = \frac{N}{D} .$$

If M families are sampled, the joint likelihood of all families is

$$L(\theta, \pi) = \prod_{m=1}^{M} L_m (\theta, \pi) ,$$

where $L_m(\theta, \pi)$ denotes the likelihood of the mth sampled family ($m = 1, 2, \ldots, M$). Since there are two parameters to be estimated, the joint likelihood $L(\theta, \pi)$ may be numerically maximized to yield maximum-likelihood estimates of θ and π.

We note that if $\pi \approx 0$ (single selection), then $\alpha_r = 1 - (1 - \pi)^r \approx r\pi$. Hence,

$$L_m(\theta, \pi) = L_m(\theta) = \frac{r \cdot P(n_i, r_i)}{\sum_{r=1}^{n} r \cdot \sum_{\substack{(l_1, l_2, \ldots, l_g) \\ l_i \leq n_i \\ \Sigma l_i = r}} P(n_i, l_i)} ,$$

which is independent of π.

Using this method, we have calculated the joint likelihood, for various values of θ and π, of all the 188 normal × normal families each ascertained through an affected offspring. These results are given in table 3. From this table, it is seen that for any value of π the likelihood is the maximum for $\theta = .04$. For $\theta = .04$, the likelihood values are more or less equal for all values of π, although numerically it is maximum for $\pi = .05$. (For values of $\pi > .3$, ln$L$ decreases monotonically for any given value of θ; $.01 \leq \theta \leq .1$.) Since the maximum-likelihood estimate of π is small, one can assume single selection and treat the likelihood as a function of a single parameter θ, as has been described above. Assuming single selection, we find that $\hat{\theta} = .039$ and that the ln$L(\theta)$ value at $\theta = .039$ is -157.5022. The standard error of $\hat{\theta}$ is .007. Comparing the estimated value of θ with those given in table 2, we see that the data are compatible with a four-locus multiple recessive homozygosis model.

Discussion

It is a moot question whether the multiple recessive homozygosis model is the true underlying genetic

mechanism for vitiligo. This model is put forward as tentative. It surely captures the essential characteristics of the family data and gives an adequate fit to the entire data set. In normal × affected families ascertained through the affected parent, it may be noted that the segregation probability is independent of the number of loci. Therefore, no assumption had to be made regarding the relationships of gene frequencies at the various underlying loci. The agreement between the expected and the observed segregation ratios in these families is surprisingly good given that neither the prevalence rate nor the age-at-onset table (table 1) is free from sampling errors.

The analysis of data of normal × normal families ascertained through an affected offspring involved the assumption of equality of gene frequencies at the loci determining the disorder. While this assumption is arbitrary, it may be noted that (1) in the absence of a priori information, any other assumption would have been equally arbitrary and (2) setting the frequency of the disorder gene at one or more loci at values higher than those at the remaining loci increases the segregation probability because, in effect, this is tantamount to suggesting the involvement of a smaller number of loci in the determination of the disorder. Further, for computing the likelihood we have used a weighted average of the segregation probabilities from various mating types, to reduce computational time and complexity. A more appropriate method would have been to compute the overall likelihood as the weighted sum of conditional likelihoods, conditional on the mating types, the weighting factors being the mating frequencies. In any case, whether four loci are actually involved cannot be confirmed through statistical studies, although the four-locus multiple recessive homozygosis model yields predictions that may be tested in other family data sets.

Our finding that a model involving multiple unlinked loci gives an adequate fit to the vitiligo data is in agreement with the findings of the studies on genetics of pigmentation in the laboratory mouse. Quevado et al. (1987) have concluded that several genes at several loci are involved in the determination of pigmentation in the mouse. For vitiligo in man, there is also suggestive evidence of the involvement of several loci. Vitiligo has been found to be significantly associated with several autosomal marker loci—RH on chromosome 1 (Das et al. 1985b), ACP1 on chromosome 2 (Das et al. 1985b), MN on chromosome 4 (Wasfi et al. 1980), and some HLA

antigens on chromosome 6 (Metzker et al. 1980). It may further be noted that recent observations suggest that vitiligo is an autoimmune disorder with specific antibodies to melanocytes (Naughton et al. 1986). Studies in mice on insulin-dependent diabetes—an autoimmune disease—have shown that three recessive loci are involved in the determination of the disease (Prochazka et al. 1987). It may well be that vitiligo is similarly determined by multiple recessive loci. Our statistical findings are in concordance with known observations.

Acknowledgments

References

Behl, P. N. 1955. Leucoderma and its treatment with Ammi Majus. J. Indian Med. Assoc. 24:615–618.

Carnevale, A., C. Zavala, V. D. Castillo, R. R. Maldonado, and L. Tamayo. 1980. Análisis genetico de 127 familias con vitiligo. Rev. Invest. Clin. 32:37–41.

Cockayne, E. A. 1933. Inherited abnormalities of the skin and appendages. Oxford University Press, London.

Das, S. K., P. P. Majumder, R. Chakraborty, T. K. Majumdar, and B. Haldar. 1985a. Studies on vitiligo. I. Epidemiological profile in Calcutta, India. Genet. Epidemiol. 2:71–78.

Das, S. K., P. P. Majumder, T. K. Majumdar, and B. Haldar. 1985b. Studies on vitiligo. II. Familial aggregation and genetics. Genet. Epidemiol. 2:255–262.

Elandt-Johnson, R. C. 1971. Probability models and statistical methods in genetics. John Wiley & Sons, New York.

El-Mofty, A. M. 1968. Vitiligo and psoralens. Pergamon, New York.

Goudie, R. B., D. R. Goudie, H. M. Dick, and M. A. Ferguson-Smith. 1980. Unstable mutations in vitiligo, organ-specific autoimmune diseases, and multiple endocrine adenoma/peptic ulcer syndrome. Lancet 2:285–287.

Hafez, M., L. Sharaf, and S. M. A. El-Nabi. 1983. The genetics of vitiligo. Acta Dermatovener. (Stockh.) 63:249–251.

Howitz, J., H. Brodthagen, M. Schwartz, and K. Thomsen. 1977. Prevalence of vitiligo: epidemiological survey of the isle of Bornholm, Denmark. Arch. Dermatol. 113:47–52.

Indian Chemical Society. 1984. Vitiligo: a major national malady (editorial). P. 1 in Indian Chemical Society, ed. Proceedings of the National Symposium on Pigment Metabolism in Relation to Vitiligo. Indian Chemical Society, Calcutta.

Lerner, A. B. 1959. Vitiligo. J. Invest. Dermatol. 32:285–310.

Levai, M. 1958. A study of certain contributory factors in the development of vitiligo in south Indian patients. Arch. Dermatol. 78:364–371.

Li, C. C. 1987. A genetical model for emergenesis: in memory of Laurence H. Synder, 1901–86. Am. J. Hum. Genet. 41:517–523.

Mehta, N. R., K. C. Shah, C. Theodore, V. P. Vyas, and A. B. Patel. 1973. Epidemiological study of vitiligo in Surat area. Indian J. Med. Res. 61:145–154.

Merelender, J., and J. Rywlin. 1940. Heredity of acquired vitiligo (vitiligo in three generations). Acta Dermatovener. (Stockh.) 21:583–585.

Metzker, A., R. Zamir, E. Gazit, M. David, and E. J. Feuerman. 1980. Vitiligo and the HLA system. Dermatologica 160:100–105.

Mohr, J. 1951. Vitiligo in a pair of monovular twins. Acta Genet. Stat. Med. 2:252–255.

Mosher, D. B., T. B. Fitzpatrick, and J. P. Ortonne. 1979. Abnormalities of pigmentation. Pp. 582–590 in T. B. Fitzpatrick, A. Z. Eisen, K. Wolf, I. K. Freedberg, and K. F. Austen, eds. Dermatology in general medicine. McGraw-Hill, New York.

Naughton, G. K., D. Reggiardo, and J.-C. Bystryn. 1986. Correlation between vitiligo antibodies and extent of depigmentation in vitiligo. J. Am. Acad. Dermatol. 15:978–981.

Prochazka, M., E. H. Leiter, D. V. Serreze, and D. L. Coleman. 1987. Three recessive loci required for insulin-dependent diabetes in nonobese diabetic mice. Science 237:286–289.

Quevedo, W. C., T. B. Fitzpatrick, G. Szabo, and K. Jimbow. 1987. Biology of melanocytes. Pp. 224–251 in T. B. Fitzpatrick, A. Z. Eisen, K. Wolff, I. M. Freedberg, and K. F. Austen, eds. Dermatology in general medicine. Vol. 1. McGraw-Hill, New York.

Siemens, H. W. 1953. Die Zwillingspathologie der Vitiligo. Acta Genet. Med. Gemellol. (Roma) 2:118–124.

Wasfi, A. I., N. Saha, H. A. El-Munshid, F. S. El-Sheik, and M. A. Ahmed. 1980. Genetic association in vitiligo. Clin. Genet. 17:415–417.