J. Genet., Vol. 74, Numbers 1 & 2, April and August 1995, pp. 41–46. © Printed in India.

Pedigree analysis of vitiligo: further support for multilocus involvement

SWAPAN K. NATH*, JAMES J. NORDLUND† and PARTHA P. MAJUMDER*

*Anthropometry and Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India
†Department of Dermatology, University of Cincinnati Medical Center, Cincinnati, OH 45267, USA

MS received 1 August 1995

Abstract. Vitiligo is a dermatological disorder in man that shows familial aggregation. We performed segregation analysis on data pertaining to vitiligo on members of 147 pedigrees each ascertained through a single proband, and tested various non-genetic, and one-locus and two-locus genetic models. Non-genetic and one-locus genetic models were rejected in favour of a two-locus model postulating epistatic interaction of recessive alleles in the aetiology of vitiligo. The present results show that vitiligo is not a single-locus disorder and substantiate our earlier inference, drawn on the basis of nuclear-family data, of multilocus involvement in the pathogenesis of vitiligo.

Keywords. Vitiligo; multilocus recessivity; segregation analysis.

1. Introduction

Genetic analysis of complex human traits has gained considerable importance in recent times (Lander and Schork 1994). Such traits usually have variable age at onset, and while they show significant familial aggregation they do not segregate in a clear-cut Mendelian fashion. Traditional analyses of family data on such traits invoked the polygenic model; however, greater attention is now being paid to Mendelian models with a small number of major loci.

In a recent paper (Nath *et al.* 1994) we have cross-validated a multilocus genetic model for vitiligo, a hypopigmentary dermatological disorder, that was proposed earlier by Majumder *et al.* (1988). For the purpose of the cross-validation study, fresh family data were gathered and data on nuclear families were analysed. Details of methodology of data collection and genetic analyses are provided in Nath *et al.* (1994). For brevity, we summarize below some salient epidemiological features of vitiligo and the major findings of our earlier study. The prevalence of vitiligo varies between 0·5% and 1% in different populations (Howitz *et al.* 1977; Das *et al.* 1985). The age at onset is variable. Mean age at onset is about 20 years. There are no significant gender differences in age-at-onset distributions. Familial aggregation, as measured by relative risks, is highly significant. The pattern of relative risks for different categories of relatives of the probands is not consistent with a single-locus mode of inheritance. The genetic model that was proposed (Majumder *et al.* 1988) and cross-validated (Nath *et al.* 1994) is that vitiligo is controlled by

a set of unlinked, diallelic loci (*A*, *a*; *B*, *b*; *C*, *c*; . . .). Individuals homozygous at each of these loci for the recessive alleles (*a*, *b*, *c*, . . .) are affected; individuals of the remaining genotypes are normal. Thus, of the many possible genotypes, individuals of one genotype (*aabbcc*. . .) are affected. The best estimate of the number of the loci was found to be three, based on comparisons of values of the likelihood function for the nuclear-family data for different numbers of loci.

While our previous study focussed on nuclear families of the 147 affected probands (Caucasian, resident in the USA), data were actually collected on extended families of the 147 probands, of which 67 (45·58%) were three-generational, 57 (38·77%) were four-generational and 23 (15·65%) were five-generational families. The analyses of data on nuclear families involved a total of 674 individuals of whom 165 were affected. The present analysis of data on extended families involved 2256 individuals of whom 216 were affected. In the present paper, we have performed a likelihood analysis of the data on extended families with a view to further verifying the validity of our finding of involvement of multiple loci in the pathogenesis of vitiligo. We emphasize at the outset that because of the enormous computational complexity involved in the evaluation of the likelihoods of data on extended families under multilocus models, the objective of the study was restricted to testing the hypothesis that two-locus recessivity provides a better fit to the data than one-locus recessivity; estimation of the number of loci was not possible. The affirmative inference of the present study with respect to the hypothesis stated above bolsters our previous finding (Majumder *et al.* 1988; Nath *et al.* 1994) that vitiligo is not a single-locus recessive disorder.

2. Methods and models

Segregation analyses of the data on members of 147 pedigrees each ascertained through an affected proband was performed using the Pedigree Analysis Package (PAP), revision 3.0 (Hasstedt 1989), which uses the nonlinear optimization routine GEMINI (Lalouel 1979). PAP uses the 'peeling algorithm' (Cannings *et al.* 1976, 1978), a generalization of the Elston–Stewart algorithm (Elston and Stewart 1971), to compute likelihoods of pedigrees. PAP incorporates an ascertainment bias correction by dividing the likelihood of the pedigree by the likelihood of the proband(s). A logistic distribution function, which has the form

$$F(x) = \{1 + \exp[-(x-a)/k]\}^{-1},$$

where $k = \sqrt{3}\, b/\pi$ or $b = k\pi/\sqrt{3}$, and a is the mean and b the standard deviation, was used as an approximation to the cumulative age-specific prevalence of the disorder. We have examined whether the logistic distribution fits the observed cumulative age-at-onset distribution by using the Kolmogorov–Smirnov nonparametric test statistic (Chakravarti *et al.* 1967). This test statistic is

$$d = \sqrt{n} \sup_x |F(x) - \hat{F}(x)|,$$

where n denotes the sample size, $F(x)$ is the logistic distribution function as defined above, and $\hat{F}(x)$ is the observed cumulative age-at-onset distribution. Under the null hypothesis that the two distributions are equal, d follows an $N(0, 1)$ distribution.

A threshold model postulating that an individual of a specific genotype and age is affected when her/his liability exceeds a certain threshold was assumed. Likelihoods, corrected for ascertainment bias, were computed for non-genetic, and various one-locus and two-locus recessive genetic models.

Before considering specific genetic models it was of interest to examine how well the data can be explained without invoking any genetic basis for the disorder, that is assuming that the disorder is solely due to environmental causes. To examine this we have considered a non-genetic/sporadic model, where the frequency of the allele causing the disorder (q) was set at 1·0, so that all individuals were of the same genotype. The prevalence of the disorder, which is the conditional probability of being affected given the genotype, was set at a fixed value (= 0·005 and 0·01) or was estimated. These three cases are denoted by N_1, N_2 and N_3. We have considered three one-locus models (R, D and G) and two two-locus models (R_1 and R_2). For one-locus models, we assumed that the locus is diallelic and designated the allele causing the disorder as a and the alternative allele as A; the frequency of the a allele in the population as q; and lifetime prevalences of the genotypes as l_1, l_2, etc. We have assumed Hardy–Weinberg equilibrium and Mendelian transmission. Model R is the recessive model with complete penetrance. In this model we fixed the lifetime prevalences of the genotypes as $l_1(aa) = 1$ and $l_2(Aa) = l_3(AA) = 0$. The only parameter of this model is q.

Model D is the dominant model with complete penetrance. In this model we fixed $l_1(aa) = l_2(Aa) = 1$ and $l_3(AA) = 0$. This model also has only one parameter, q.

Model G is a general genetic model which allows for both incomplete penetrance in the susceptible genotype (aa) and the presence of phenocopies among individuals of the nonsusceptible genotypes (AA, Aa). This model, therefore, has four parameters q, l_1, l_2 and l_3.

For the two-locus model, we considered two unlinked diallelic loci. We designated the alleles at the two loci as a and b and their counterparts as A and B, respectively. Under the postulated two-locus recessive model, out of the ten possible two-locus genotypes, individuals of only one genotype ($aabb$) are affected; individuals of the other nine genotypes are normal. Under this model, there are ten genotypic lifetime prevalences, denoted by $l_1, l_2, \ldots, l_{10}$, of which l_1 corresponds to the genotype $aabb$, and $l_2, l_3, \ldots, l_{10}$ to the remaining nine genotypes. For the present two-locus recessive model under consideration, $l_1 = 1$ and $l_2 = \ldots = l_{10} = 0$. These values were held fixed in all likelihood computations. The frequencies of alleles a and b were held equal (denoted by q) since the separate estimation of the two allele frequencies is mathematically unfeasible. This is because allele frequencies are estimated from the proportion of affected individuals among founders (including individuals marrying into the pedigree); this proportion is equal to, except for sampling fluctuations, the population prevalence, which is a function solely of the product of frequencies of alleles a and b. Likelihood computations were performed in two ways: model R_1, in which q was held fixed at 0·2659 (which corresponds to a population prevalence of 0·5%); and model R_2, treating q as an independent parameter to be estimated from the data.

For the purpose of cross-checking of estimates and inferences and also for computational convenience, the set of 147 pedigrees was randomly split into two approximately equal subsets (subset I and subset II), comprising 75 and 72 pedigrees

including 1185 and 1071 individuals, respectively. All computations were performed separately for the two subsets.

3. Results and discussion

The general form of the observed cumulative age-at-onset distribution appears to be that of the logistic distribution function. In figure 1 we have plotted the observed cumulative age-at-onset distribution and the logistic distribution function with parameters $a = 20$ and $b = 15$. (These parameter values are approximately equal to the observed mean and standard deviation of the ages at onset among probands.) For these parameter values of a and b, the observed value of the Kolmogorov–Smirnov test statistic d was 1·32, which is non-significant at the 5% level, implying that the logistic approximation to the observed age-at-onset distribution is acceptable.

The results of model fitting (segregation analysis) are presented in table 1, separately for the two subsets. As mentioned earlier, the three non-genetic models N_1, N_2 and N_3 correspond, respectively, to situations in which the population prevalence of vitiligo was held fixed at 0·005 or 0·01 or was estimated from the data. For all non-genetic models, the frequency of the allele causing the disorder was held fixed at 1. It is seen that the likelihood of the data under the non-genetic model increases with increase in prevalence and the maximum likelihood estimate of prevalence obtained for N_3 is about 4% from both subsets of the data.

Among the one-locus models considered, the dominant model (D) is clearly rejected; compared to other models the $-2\ln(LR)$ (LR = likelihood ratio) varies between 1268·9 and 1404·7 for subset I, and between 1223·7 and 1328·1 for subset II. The likelihoods under the recessive model (R) and the non-genetic model N_3 are of a similar magnitude, but the estimate of prevalence (0·005), derived from the maximum likelihood estimate of the allele frequency, for the recessive model is much closer to observed population prevalences (0·005–0·01) than the estimate of 0·04 obtained under the non-genetic model. The general one-locus model (G) provides the best fit to the data and is significantly better than the one-locus

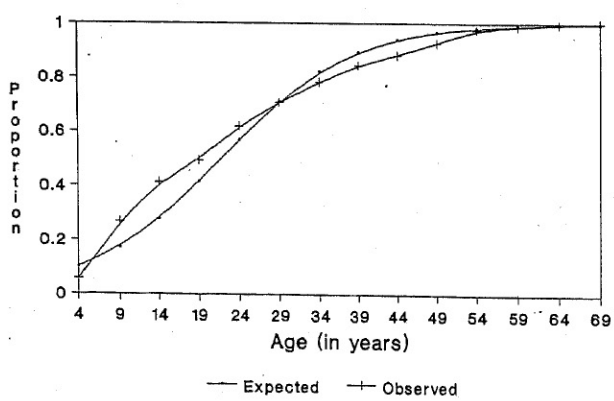

Figure 1. Cumulative distribution of age-specific prevalences of vitiligo: observed and expected distributions. (Note: The expected distribution was obtained assuming a logistic distribution function with mean = 20 and standard deviation = 15.)

Table 1. Results of pedigree analysis of vitiligo for non-genetic, and one-locus and two-locus genetic models.

Data subset (*N*=No. of families)	Parameter	Non-genetic models			Genetic models: One-locus models			Genetic models: Two-locus models	
		N_1	N_2	N_3	R	D	G	R_1	R_2
Subset I ($N=75$)	q	[1]	[1]	[1]	0·0680	0·0500	0·0230	[0·2659]	0·2415
	l_1	—	—	—	[1]	[1]	0·9975	[1][b]	[1][b]
	l_2	—	—	—	[0]	[1]	0·0787	—	—
	l_3				[0]	[0]	0·0046	—	—
	Prevalence[a]	[0·005]	[0·01]	0·038	0·005	0·098	0·010	0·005	0·003
	$\log_{10} L$	−96·31	−87·22	−77·44	−77·48	−371·91	−69·12	−68·04	−66·87
	$-2 \ln L$	443·83	401·68	356·64	356·84	1712·7	318·31	313·33	307·97
Subset II ($N=72$)	q	[1]	[1]	[1]	0·0689	0·0400	0·0525	[0·2659]	0·2485
	l_1	—	—	—	[1]	[1]	0·6491	[1][b]	[1][b]
	l_2	—	—	—	[0]	[1]	0·0597	—	—
	l_3	—	—	—	[0]	[0]	0·0108	—	—
	Prevalence[a]	[0·005]	[0·01]	0·038	0·005	0·078	0·017	0·005	0·004
	$\log_{10} L$	−85·57	−77·46	−68·70	−70·62	−351·37	−64·73	−63·67	−62·88
	$-2 \ln L$	394·05	356·70	316·37	325·20	1617·7	298·09	293·22	289·56

Figures in brackets are fixed values of parameters; those not in brackets are maximum likelihood estimates of parameters.

a, Prevalence was calculated from estimates of parameters.

b, l_1 refers to lifetime prevalence of genotype *aabb*; the lifetime prevalences of the remaining nine genotypes, l_2, l_3, . . . , l_{10}, were fixed at 0.

recessive model ($-2 \ln(\text{LR}) = 38{\cdot}5$ for subset I and 27·1 for subset II; d.f. = 3; $p < 0{\cdot}05$ for both subsets). For both subsets, estimated prevalences under the general model are higher than those obtained under the recessive model, but nevertheless are close to the observed population prevalences reported earlier. A careful examination of the estimated lifetime prevalences of the three genotypes obtained under the general one-locus model indicates that these estimates are similar, in particular for subset I, to the lifetime prevalences of the genotypes of the recessive model. In other words, the general model is effectively the same as the recessive model.

When compared with the best-fitting one-locus model G, the two-locus model is found to provide a substantially better fit ($-2 \ln(\text{LR}) \approx 10$). Because of differences in parametric structures of the one-locus and two-locus classes of models, we were unable to perform rigorous statistical tests of significance to compare these two classes of models. Further, because of computational complexities it was not possible to perform likelihood computations on pedigree data for genetic models involving more than two loci. Because of this limitation, the present pedigree analysis only confirms whether a multilocus model provides a better fit than a single-locus model, but does not provide an estimate of the number of loci involved. We were thus unable to verify our previous estimate that three loci are involved in the pathogenesis of vitiligo. Our present analyses, however, show that the pedigree data on vitiligo fit a two-locus recessive model better than a one-locus recessive model; in other

words, vitiligo is not a single-locus disorder but is controlled by multiple loci. It is also interesting to note that the population prevalences calculated from the estimates of parameters obtained under the two-locus recessive model are the same ($\approx 0{\cdot}004$) for both subsets of the data. This value of estimated prevalence is also very close to earlier population estimates. In conclusion, the present results substantiate our previous finding of epistatic interaction of recessive alleles at multiple unlinked loci in the pathogenesis of vitiligo. This inference makes considerable biological sense. It is known that there are several key points in the human pigmentary pathway (Hearing and King 1993). The implicated loci may serve as controls of these key points and recessive homozygosis at any of these loci may be viewed as complete disruption or blockage at the corresponding point. If one of these control points is blocked then bypass routes in the pigmentary pathway are possibly used and there is no precipitation of disorder state. However, if a greater number of control points are blocked, then possibly there is failure of the entire pathway because of non-availability of bypass routes, which results in a disruption of the end product or process and consequently in the clinical manifestation of vitiligo.

We finally note that the model proposed here has also been found to provide possible explanations for a number of puzzling observations in several diseases, such as Cockayne's syndrome and Fanconi's anaemia, associated with defective DNA repair (Lambert and Lambert 1992).

References

Cannings C. A., Skolnick M. H., De Nevers K. and Sridharan R. 1976 Calculation of risk factors and likelihoods for familial diseases. *Comp. Biomed. Res.* 9: 393–398

Cannings C. A., Thompson E. A. and Skolnick M. 1978 Probability functions on complex pedigrees. *Adv. Appl. Prob.* 10: 26–61

Chakravarti I. M., Laha R. G. and Roy J. 1967 *Handbook of methods of applied statistics, Volume I: Techniques of computation, descriptive methods and statistical inference* (New York: John Wiley & Sons)

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

Elston R. C. and Stewart J. 1971 A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21: 523–542

Hasstedt S. J. 1989 Pedigree Analysis Package (revision 3.0). University of Utah, Salt Lake City, Utah, USA

Hearing V. J. and King R. A. 1993 Determinants of skin color: melanocytes and melanization. In *Pigmentation and pigmentary disorders* (ed.) N. Levin (Boca Raton: CRC Press, Inc.) pp. 3–32

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

Lalouel J. M. 1979 GEMINI: A computer program for optimization of general non-linear functions. Technical report 14, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City, Utah, USA

Lambert W. C. and Lambert M. W. 1992 Co-recessive inheritance: A model for DNA repair and other surveillance genes in higher eukaryotes. *Mut. Res. (DNA Repair)* 273: 179–192

Lander E. S. and Schork N. J. 1994 Genetic dissection of complex traits. *Science* 265: 2037–2048

Majumder P. P., Das S. K. and Li C. C. 1988 A genetical model for vitiligo. *Am. J. Hum. Genet.* 43: 119–125

Nath S. K., Majumder P. P. and Nordlund J. J. 1994 Genetic epidemiology of vitiligo: Multilocus recessivity cross-validated. *Am. J. Hum. Genet.* 55: 981–990