Am. J. Hum. Genet. 52:1006–1011, 1993

Nonsyndromic Cleft Lip With or Without Cleft Palate in West Bengal, India: Evidence for an Autosomal Major Locus

Ajit K. Ray,* L. Leigh Field,† and Mary L. Marazita‡

*University of Toronto, Toronto; †University of Calgary and Alberta Children's Hospital, Calgary; and ‡Medical College of Virginia, Richmond

Summary

Ninety extended families having one or more individuals affected with nonsyndromic cleft lip (CL) with or without cleft palate (CL/P) were ascertained in rural West Bengal, India. These families included 138 affected people, 64% of whom had CL alone and 66% of whom were male. Multiple-affected-member ("multiplex") pedigrees were less common than single-affected-member ("simplex") pedigrees, composing 34% of all extended pedigrees. There was no difference between multiplex and simplex pedigrees in the frequency of affected persons with CL alone, but multiplex pedigrees had a lower frequency of affected males (58%) than did simplex pedigrees (76%; $P = .02$). Complex segregation analysis using the POINTER computer program rejected both the hypothesis of no familial transmission ($P < .0001$) and the hypothesis that familiality could be explained solely by a multifactorial/threshold model ($P < .05$). The hypothesis of major-locus inheritance alone could *not* be rejected. Among major-locus models examined, strictly recessive inheritance was rejected ($P < .0001$), but codominant and dominant models were not. Neither the addition of a multifactorial component nor the addition of a proportion of sporadic cases to the major-locus model improved the fit of the data. In conclusion, the results of complex segregation analysis were consistent with a dominant or codominant major-locus mode of inheritance of CL/P in these families.

Introduction

The nature of the genetic contribution to the etiology of nonsyndromic cleft lip (CL) with or without cleft palate (CL/P) remains controversial, with some investigators apparently continuing to support the earlier model of multifactorial inheritance with a threshold (MFT; polygenic; e.g., see Mitchell and Risch 1992) and others espousing models involving a more limited number of genetic loci (major-locus or oligogenic models) (see the recent review in Marazita et al. 1992; also see Farrall and Holder 1992). At times, the debate has seemed to be primarily about semantics, particularly since the terms "polygenic" and "oligogenic" may overlap (depending on one's perception of the meaning of "many" vs. "few"), and, similarly, "oligogenic" and "major"/"single" locus may overlap (depending on whether one's definition of a major-locus model allows for a few nonmajor modifying loci). To this plethora of terminology must be added the possibility of genetic heterogeneity—i.e., multiple major loci which independently can cause the disorder. Note that the frequencies of such multiple major loci, if they exist, may vary considerably from population to population.

However, with the advent of numerous highly polymorphic genetic markers which can be used to locate major disease loci through linkage analysis, the question of the mode of inheritance of CL/P is no longer simply a debate about semantics. If there is a major locus which is necessary to the development of CL/P in a large proportion of families with multiple affected members, then it should be possible to locate that major locus in those families by using linkage analysis. On the other hand, if CL/P is determined by the accumulative minor effects of many loci, or if there are several independent major loci (a high degree of genetic heterogeneity), then it may not be possible to locate these predisposing loci through linkage analysis of a "reasonable" number of families. In this paper, we present evidence that a major locus contributing to CL/P is detectable by complex segregation analysis in 90 extended families living in West Bengal, India.

Received December 11, 1992.

Address for correspondence and reprints: Dr. L. Leigh Field, Department of Pediatrics, Health Sciences Centre, 3330 Hospital Drive N.W., Calgary, Alberta T2N 4N1, Canada.

0002-9297/93/5205-0019$02.00

Subjects and Methods

Study Population

Ninety extended families with nonsyndromic CL/P were ascertained during 1987–89, by one of us (A.K.R., fluent in the local languages and customs), among both Hindus and Muslims in the rural areas of the Hoogly and 24 Parganas districts near Calcutta, West Bengal, India. Children at schools and adults in the local markets were asked whether they knew of persons with CL/P. These persons were then located, and the families were interviewed with regard to family history of CL/P in both maternal and paternal lineages over a minimum of three generations, with particular emphasis on identifying other affected pedigree members and their relationship to the proband. In these rural families surveyed, the CL/P malformation usually remains uncorrected because of poor financial circumstances and lack of surgical resources; babies are born outside hospitals, and mortality due to feeding problems among infants with cleft palate is high (they cannot create suction necessary for nursing). The presence of CL/P in living individuals was verified by the fieldworker. Each extended pedigree contained only one proband—that is, there was a low ascertainment probability (incomplete single selection). Segregation analyses were therefore performed using an ascertainment probability of .01. However, duplicate analyses at a higher ascertainment probability of .30 were also performed to assess the effect of variation in that parameter.

Segregation Analyses

In order to compare the fit of the family data to multifactorial/threshold (MFT), Mendelian major-locus (ML), and combined-MFT-plus-ML models, the extended pedigrees were analyzed by complex segregation analysis using the unified mixed model (Morton and MacLean 1974; Lalouel and Morton 1981; Lalouel et al. 1983) as implemented in the computer program POINTER (Morton et al. 1983). This program only accepts nuclear families as input; extended pedigrees were analyzed by dividing them into their component nuclear families and indicating through the use of "pointers" how nuclear families not containing probands were ascertained for study—i.e., their relationship to the proband nuclear family within each extended pedigree. Dividing the 90 extended pedigrees into nuclear families resulted in 593 nuclear families being included in the segregation analyses: 117 families containing probands and possibly other affected members (note that there were more than 90 such families because some probands were included twice in the analyses, as a child in one family and as a parent in another), 36 families with affected members who were not probands, and 440 families with no affected members. The probabilities of children's phenotypes were conditioned on parental phenotypes, so that including some probands twice in the analyses did not bias the results.

Eleven pedigrees included a consanguineous mating which complicated the assignment of pointers indicating relationship of proband to nonproband families. In seven of these pedigrees, there was only one affected person, and we chose to retain only the proband nuclear family in the analyses. In the remaining four families with multiple affected persons, all nuclear families were retained, and degree of relationship to the proband was modified, when necessary, to reflect closer kinship due to consanguinity.

Likelihoods of the data under various models and maximum-likelihood estimates of relevant parameters were calculated using the POINTER program. Three of the estimated parameters are relevant to the characterization of the ML model: d, the degree of dominance; t, the displacement between the two homozygotes; and q, the gene frequency of the disease allele. The parameter relevant to the MFT model is H^2, the heritability due to additive polygenes. The final parameter, X, represents the proportion of sporadic cases of mutational or purely environmental origin. Transmission probabilities at the major locus (τ's) were not estimated but, rather, were fixed at their Mendelian values, because of a reported problem in transmission probability estimation using the POINTER program when selection is not complete or when pointers are employed (Iselius and Morton 1991). Hypothesis testing was performed by comparing the fit of specific restricted models to the general unrestricted model by using the likelihood-ratio criterion, wherein (−2 ln likelihood of the restricted model) minus (−2 ln likelihood of the general model) is distributed as χ^2, with df equal to the difference in number of parameters estimated. To determine the most parsimonious of the best-fitting models, the Akaike information criterion (AIC; Akaike 1974) was calculated for each model, as AIC = −2 ln likelihood of the model + 2(number of estimated parameters).

Results

Study Population

The distribution of CL/P by sex, type of malformation, and type of family (multiplex—"more than one

Table 1

Distribution of 138 CL/P Subjects, by Sex, Type of Malformation, and Type of Family

	CL Only	CL+P	Total
Male	59	32	91 (.66)
Female	29	18	47 (.34)
Total	88 (.64)	50 (.36)	138
	Simplex	Multiplex	
CL only	39 (.66)	49 (.62)	
CL+P	20 (.34)	30 (.38)	
Total	59	79	138
Male	45 (.76)	46 (.58)	
Female	14 (.24)	33 (.42)	
Total	59	79	138

affected in the extended pedigree"; or simplex—"one affected in the extended pedigree") is presented in table 1. Among the 90 extended pedigrees ascertained, 31 (34%) contained more than one affected individual. Of these 31 multiplex families, 17 contained two affected individuals, 11 contained three affected individuals, and 3 contained four affected individuals. Thus, in these 90 extended pedigrees, there were a total of 138 persons having CL/P. Of these, 66% (91/138) were male (a sex ratio of 1.94, close to a 2:1 ratio of male:female), and 64% (88/138) had CL alone with no palate involvement. The percentage of affected individuals who had CL alone did not differ between males and females (59/91 = 65% vs. 29/47 = 62%, respectively; $\chi^2 = 0.13$). Thus, there was no difference in the severity of affection (percentage of CL alone vs. CL and palate [CL+P]) by sex.

There was no difference between simplex and multiplex pedigrees in the percentage of affecteds with CL alone (66% vs. 62%, respectively; $\chi^2 = 0.24$). That is, there was no indication of a "severity effect" according to the number of affected individuals in the pedigree. However, there was a significant difference between simplex and multiplex pedigrees in the percentage of affected individuals who were male (76% vs. 58%, respectively; $\chi^2 = 5.44$, $P = .02$). In simplex families, the sex ratio of affected people was significantly different from 50% (76% males; $\chi^2 = 16.29$, $P < .0001$), while in multiplex pedigrees it was not (58% males; $\chi^2 = 2.14$).

There were 11 extended pedigrees with consanguineous matings, including 1 pedigree in which there were two such matings. Although we do not have access to information on the frequency of consanguineous matings in West Bengal, our impression is that the consanguineous marriage rate in this sample of CL/P pedigrees is no greater than would be found in similar-size and generation-depth pedigrees drawn at random from the population in this area. Of the 12 consanguineous matings, only 6 (50%) resulted in affected children (4 first-cousin matings, each with one affected child; 1 first-cousin mating, with two affected children; and 1 second-cousin mating, with two affected children). The percentage with CL alone among the eight affected children was the same as in the total sample (5/8 = 63%). However, this small sample of affected inbred children did *not* show a preponderance of males (three males and five females).

Segregation Analyses

The results of the complex segregation analyses on the 90 extended pedigrees are presented in table 2. When the parameters of the general unrestricted mixed model (model 1 in table 2) were estimated, the value of H^2 converged to a boundary value. Thus, only three parameters were actually estimated in this model, and the df for hypothesis testing were therefore reduced by 1. Similarly, for the MFT model (model 4), the value of H^2 also converged to a boundary value, so that, for purposes of hypothesis testing, no parameters were estimated.

The hypothesis of no familial transmission could be clearly rejected (model 2 vs. 1; $\chi^2 = 209.22$, 3 df, $P < .0001$). The hypothesis of a strictly MFT mode of inheritance with no ML could also be rejected (model 4 vs. 1; $\chi^2 = 8.52$, 3 df, $P < .05$). On the other hand, the hypothesis of strictly ML inheritance with no MFT component could *not* be rejected (model 3a vs. 1; $\chi^2 \sim 0.0$). This is readily apparent by inspection of the value of H^2 in the most general model (model 1): it converged to .0, with a likelihood almost identical to that of the ML model (model 3a). By the AIC, the ML model is more parsimonious than the general mixed model, since its AIC is smaller.

Among the ML models, the hypothesis of recessive ($d = .0$) inheritance could be rejected (model 3b vs. 3a; $\chi^2 = 48.8$, 1 df, $P < .0001$). However, codominant ($d = .5$) and dominant ($d = 1.0$) models could *not* be rejected; they produced likelihoods very similar or identical to that of the general ML model. Thus, while strictly recessive inheritance could be discounted, the precise degree of dominance at the ML could not be further assessed. By the AIC, the dominant model is the most parsimonious.

Table 2

Results of Complex Segregation Analyses of CL/P Extended Kindreds from West Bengal

Hypothesis	Value of Parameter						
	d	t	q	H^2	−2 ln Likelihood	χ^2	AIC
1. General mixed model							
ML+MFT	.42	5.72	.001	0[a]	−76.92	...	−70.92
2. No familial transmission							
(d=t=q=H^2=0)	[0]	[0]	[0]	[0]	132.30	209.22	132.30
3. ML, no MFT (H^2=0):							
a. General	.79	3.03	.001	[0]	−76.99	~0	−70.99
b. Recessive	[0]	4.18	.065	[0]	−28.19	48.8	−24.19
c. Codominant	[.5]	4.80	.001	[0]	−76.94	.05	−72.94
d. Dominant	[1.0]	2.41	.001	[0]	−76.99	0	−72.99
4. MFT, no ML (d=t=q=0)	[0]	[0]	[0]	1.0[a]	−68.40	8.52	−68.40

NOTE.—Numbers in square brackets represent parameters which were not estimated but, rather, were fixed to the value inside the brackets.
[a] Parameter converged to a boundary value.

Note that none of the models in table 2 includes the parameter X, the proportion of sporadic cases. The reason for this is that X could not be estimated in the general unrestricted model (model 1 in table 2), since H^2 converged to its boundary value of .0 before the value of X could change substantially (although it did decrease). When this parameter was estimated in a codominant ML model (model 3c in table 2), the value converged to .0, and the likelihood was identical to that obtained when X was not included in the model. We therefore concluded that there was no evidence that the addition of sporadics to the ML model improved the fit in these data.

Since the exact ascertainment probability was uncertain, we repeated all segregation analyses using a probability of .30. Results of these analyses (data not shown) were very similar to those produced by using an ascertainment probability of .01.

In summary, the results of the complex segregation analyses support the hypothesis that inheritance of CL/P in these families is determined by a *major nonrecessive autosomal locus,* with no contribution from additive polygenic (i.e., MFT) or sporadic sources. The AIC values are also consistent with this conclusion, since, of the models examined, the dominant ML model (model 3d) has the smallest AIC value (see table 2).

Discussion

The proportion of multiplex families in our data set (34%) was similar to that reported in previous large studies of CL/P (e.g., 37% in the Danish data of Fogh-Andersen 1942). However, in our data set the frequency of CL+P (50/138 = 36%) was less than that of CL alone, while in most other studies CL+P has predominated. Possible explanations for this phenomenon in our data are (1) underreporting of babies with CL+P who have died in infancy; (2) erroneous reporting of dead relatives as having cleft lip only, an error due to the hidden nature of cleft palate; and (3) a truly altered CL+P:CL ratio, as a result of generally inadequate prenatal nutrition (producing either more fetuses with CL only or increased loss of fetuses with CL+P) or as a result of differences in the frequency of genes modifying severity of expression of CL/P. Records of hospital births at Calcutta Medical College during 1980–86 (A.K.R., unpublished data) showed that 52% (11/21) of CL/P newborns had CL+P, suggesting that rural partially historical data may in fact be different from urban hospital-based data. However, this still does not enable one to determine the cause(s) of that difference. A recent study by Nemana et al. (1992) has reported that 62% of CL/P subjects undergoing surgery in a Madras hospital had CL+P. Note that surgical cases may have different CL+P:CL proportions than do unrepaired newborns and that both the Calcutta hospital newborns and Madras hospital surgical patients undoubtedly had better economic/nutritional resources than did our rural study sample.

Our data are consistent with the majority of studies which have demonstrated more affected males than females. This supports the concept that susceptibility to the defect (penetrance) is higher in males than in females, as a result of sex-specific differences in develop-

mental physiology. Degree of severity of the defect (expression) does not appear to be influenced by sex in our data set: the male:female ratio for CL alone was 2.03, and for CL+P it was 1.78 (difference not significant). In data sets from Indiana (Dronamraju et al. 1984) and China (Marazita et al. 1992), the sex ratio was also higher for CL than for CL+P, but the reverse prevailed in a large data set from Denmark (Melnick et al. 1980). It may be that *penetrance* and *severity* are influenced by different factors—and that sex influences penetrance more than it does severity.

Since the percentage of affected individuals with CL alone did not differ between simplex and multiplex pedigrees, there is no indication, in our data set, of a "severity effect" (CL+P vs. CL alone) by number of affected family members. In other words, there is no evidence of a higher recurrence risk in families with more severely affected probands. This contradicts a postulate of the traditional MFT model, which holds that, because of a higher concentration of genetic and/or environmental liability factors, recurrence risks will be higher in families with more severely affected probands. In their family data set from Madras, Nemana et al. (1992) also saw no evidence for such an effect. However, we would have expected that, even if increased genetic loading were not relevant (assuming CL/P is not polygenic), increased environmental loading would still be manifested by a higher proportion of severely affected people in multiplex families. One possible interpretation of this finding is that predisposing familial environmental factors have more effect on penetrance than on severity.

There was a significantly higher frequency of affected females in multiplex than in simplex families (42% vs. 24%), such that in multiplex families the frequency of CL/P females and CL/P males approached equality. This effect has been observed in other data sets (e.g., see Niswander et al. 1972). The simplest explanation for this is that the loading of environmental factors predisposing to CL/P is higher in multiplex families and results in a higher frequency of the less-susceptible sex (females) becoming affected. In other words, penetrance of CL/P may be influenced by effects of both environmental and sex-specific developmental factors; a higher penetrance at the major locus in multiplex families that is due to shared environmental effects would obscure the otherwise notable difference in sex-specific penetrance. It is also possible that penetrance is influenced by minor ("modifying") genetic factors, although these are probably limited in number, since the addition of a polygenic component to the major-locus model of CL/P appeared unwarranted. In addition, genetic factors (unrelated to sex) may contribute to the *severity* of the defect.

Our small sample of affected inbred children also had a high frequency of females (5/8 = 63%). Perhaps the penetrance of CL/P in these inbred individuals was increased by a minor recessive modifying gene, again obscuring the usual sex differences in liability (note, however, that the segregation analysis results rejected the hypothesis of a recessive *major* locus). Larger numbers of inbred CL/P children will be needed to better assess whether there is an inbreeding effect on sex-specific penetrances. Since the proportion of CL alone among the inbred CL/P children did not differ from that in the entire data set, there is no evidence for an inbreeding effect on severity. This reinforces the concept that penetrance and expression of CL/P are influenced by different factors.

The results of complex segregation analysis of CL/P using our data were consistent with a dominant or codominant major-locus model of inheritance, with no indication that the addition of a multifactorial or sporadic component to the major-locus model improved the fit of the data. These findings from CL/P families living in West Bengal in northeastern India are remarkably similar to those of the only other complex segregation analysis on a CL/P data set from India, that of Nemana et al. (1992) studying families from Madras in southern India. They also found that the best-fitting model was a major locus with no multifactorial component and no sporadics. With transmission probabilities fixed at Mendelian values and with H^2 fixed at .0 to exclude a multifactorial component (the equivalent of our model 3a), their data set also estimated degree of dominance at the major locus as intermediate (.44).

In conclusion, complex segregation analysis of 90 extended families from rural West Bengal, India, ascertained through probands affected with CL/P has produced evidence for a major autosomal locus controlling this common birth defect. These findings suggest that it should be possible to locate the major CL/P locus in this population by analyzing genetic linkage between the CL/P trait and markers at candidate or random (genome screen) loci.

Acknowledgments

Dr. Field is an Alberta Heritage Medical Scientist. The financial support of University of Toronto SSHR grant to A.K.R., Medical Research Council of Canada grant MT-11263 to L.L.F. and A.K.R., National Institutes of Health

grant DE-07360 to M.L.M., and Alberta Heritage Foundation for Medical Research salary award to L.L.F. is gratefully acknowledged. A special thanks goes to the many CL/P families who generously contributed their interest and energies to this research project.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automatic Control AC-19:716–723

Dronamraju KR, Wakim KG, Smith DJ, Bixler D (1984) Fetal mortality in oral cleft families (IX): factors relating to the occurrence of sporadic clefts. Clin Genet 26:322–330

Farrall M, Holder S (1992) Familial recurrence-pattern analysis of cleft lip with or without cleft palate. Am J Hum Genet 50:270–277

Fogh-Andersen P (1942) Inheritance of harelip and cleft palate. A Busck, Copenhagen

Iselius L, Morton NE (1991) Transmission probabilities are not correctly implemented in the computer program POINTER. Am J Hum Genet 49:459

Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. Hum Hered 31:312–321

Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. Am J Hum Genet 35:816–826

Marazita ML, Hu D-N, Spence MA, Liu Y-E, Melnick M (1992) Cleft lip with or without cleft palate in Shanghai, China: evidence for an autosomal major locus. Am J Hum Genet 51:648–653

Melnick M, Bixler D, Fogh-Andersen P, Conneally PM (1980) Cleft lip-cleft palate: an overview of the literature and an analysis of Danish cases born between 1941 and 1968. Am J Med Genet 6:83–97

Mitchell LE, Risch N (1992) Mode of inheritance of nonsyndromic cleft lip with or without cleft palate: a reanalysis. Am J Hum Genet 51:323–332

Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. Am J Hum Genet 26:489–503

Morton NE, Rao DC, Lalouel JM (1983) Methods in genetic epidemiology. Karger, New York

Nemana LJ, Marazita ML, Melnick M (1992) A genetic analysis of cleft lip with or without cleft palate in Madras, India. Am J Med Genet 42:5–10

Niswander JD, Chung CS, MacLean CJ, Dronamraju KR (1972) Sex ratio and cleft lip with or without cleft palate. Lancet 2:858–860