COMPLEX SEGREGATION ANALYSIS OF NONSYNDROMIC CLEFT LIP/PALATE IN ANTIOQUIA, COLOMBIA

ANÁLISIS DE SEGREGACIÓN COMPLEJA DE LABIO/PALADAR HENDIDO NO SINDRÓMICO EN ANTIOQUIA, COLOMBIA

Lina Moreno¹, Andrew Lidral¹, Idalyd Fonseca², Consuelo Valencia and Mauricio Arcos-Burgos²

Abstract

The present study was undertaken to examine the pattern of inheritance of Cleft Lip/Palate (CLP) in pedigrees ascertained from Antioquia, Colombia. Ninety-five extended and multigenerational pedigrees, constituted by 201 nuclear components and 1.136 records were analyzed. Ten hypothetical models were contrasted using likelihood ratio tests. The hypotheses of no familial transmission, multifactorial component compared against that of the existence of a major gene only, the existence of a recessive major gene, that of non major component in the mixed model and that of the non transmission of major effect (t₁ = t₂ = t₃) were rejected. In contrast, hypotheses postulating a major locus (dominant, codominant) and that of no polygenic component in the mixed model could not be rejected. Iteration of the parameter t₂ was the most parsimonious. Thus far, the most parsimonious model is that of a major gene (dominant-codominant) without multifactorial effects but, taking into account, that the t₂ iteration in the major gene model with unrestricted d, result in a significantly improving of the model likelihood, oligogenic interactions can not be underrated.

Key words: Colombia, South America, cleft lip and palate, genetics, major gene, Antioquia, racial admixture, ethnic, complex segregation analysis, CLP.

Resumen

El presente estudio fue realizado para evaluar el patrón hereditario de Labio/Paladar Hendido (CLP) en genealogías estudiadas en el departamento de Antioquia, Colombia. Se analizaron 95 genealogías multigeneracionales extendidas, constituidas por 201 componentes nucleares y 1.136 individuos. Se contrastaron diez modelos hipotéticos mediante el test de verosimilitud. Las hipótesis de no-transmisión familiar y el componente multifactorial, comparadas contra la hipótesis de existencia de un gen mayor y de un gen mayor reesivo, y la no existencia del componente mayor en el modelo mixto y no transmisión de efectos mayores (t₁ = t₂ = t₃), fueron rechazadas. En contraste, no pudieron rechazarse las hipótesis que postulan un locus mayor (dominante, codominante) y un componente no poligénico en el modelo mixto. El modelo más parsimonioso fue el de gen mayor (dominante-codominante) sin efectos multifactoriales tomando en cuenta la iteración del parámetro t₂ con d no restringido, lo que resultó en un mejoramiento significativo del valor de verosimilitud en este modelo. Las interacciones oligogénicas no pueden ser subestimadas.

Palabras clave: Colombia, Suramérica, labio y paladar hendido, genética, gen mayor, Antioquia, mezcla racial, étnico, análisis de segregación compleja, CLP.

INTRODUCTION

Although the genetic component in the etiology of nonsyndromic cleft lip with or without cleft palate (CLP) has been recognized, the inheritance model explaining available data remains controversial. The most widely accepted model for CLP during the 70’s and early 80’s was that of a multi-

---

¹ Dentistry School, University of Ohio, USA. E-mail: Moreno.25@osu.edu.
² Population Genetics and Genetic Epidemiology Group, Department of Biology, University of Antioquia, Medellín. E-mail: omarcos@matematicas.udea.edu.co.
factorial inheritance (Fogh-Andersen, 1942; Carter, 1969; Fraser, 1970; Carter, 1976; Carter et al., 1982). More recently, the complex segregation analysis based on the mixed model (Lalouel and Morton, 1981; Lalouel et al., 1983) has suggested the existence of a mendelian component. Marazita et al. (1986) reexamined the data of Carter et al. (1982) and found that the multifactorial threshold model, as an unique explanation of the familial clustering of CLP, could be rejected in favor of a mixed model (single major locus plus multifactorial components). Similar results were previously reported by Demenais et al. (1984) in France. In this way, in the last ten years, considerable work on different populations, performed to elucidate the genetic model underlying susceptibility to CLP, have been suggestive of the existence of major genes. For instance, Melnick et al. in China (1986), Chung et al. (1986) on the Danish data, Bixler et al. (1971), Melnick et. al. (1980), Hecht et al. (1991) on US families, Marazita et al. (1992) in China, Nemana et al. (1992) in India, Ray et al. (1993) in West Bengal, Palomino et al. (1997) and Blanco et al. (1998a) on the Chilean population, all suggested the possibility of a major gene for clefting.

The fundamental problem with the major gene component is that dominance appears to be heterogeneous among different populations. Melnick et al. (1986) Marazita et al. (1992) and Chung et al. (1986) have found compatibility with the recessive model of inheritance, while De Paepe (1989), Temple et al. (1989), Hecht (1990), Hecht et al. (1991a), Nemana et al. (1992), Ray et al. (1993), Clementi et al. (1995), Palomino et al. (1997) and Blanco et al. (1998a) with that of the codominant-dominant model. In all cases, penetrance has been found to be incomplete. Indeed, several analysis have postulated the existence of at least two major epistatically interacting loci (Farral and Holder, 1992; Mitchell and Risch, 1992; Clementi et al., 1995).

In accordance with the hypothesis of a major gene, several studies have been carried out where various candidate genes have been analyzed in different populations, either through association or linkage studies (Artinger et al., 1989; Hecht et al., 1991b; Chenevix-Trench et al., 1991, 1992; Holder et al., 1992; Stoll et al., 1992; Vintiner et al., 1992, 1993; Sassani et al., 1993; Hecht et al., 1993; Davies et al., 1995; Carinci et al., 1995; Stein et al., 1995; Mitchell et al., 1995; Jara et al., 1995; Lidral et al., 1997, 1998; Blanco et al., 1998b). The results of these investigations have also shown inconsistent results, which seem to reflect genetic heterogeneity. Ethnical background influence and a great inter-population variability suggest that more than one susceptibility locus is involved in the etiology of CLP.

Latin American tropical countries offer an interesting situation for studying the genetics of diseases because the mixed ethnical background which originated, about ten to twenty generations ago among ameriesiandians, europeans and negroids. This would allow the use of linkage disequilibrium methods such as Mapping by Admixture Linkage Disequilibrium (MALD) as a powerful strategy to find out major genes predisposing the development of complex disease such as CLP (Chakraborty and Weiss, 1988; Blanco and Rosales, 1988; Blanco et al., 1998a; Arcos-Burgos et al., 1999).

The human population in Colombia (12° 30' 40" N, 4° 13' 30" S) stems mostly from the admixture of europeans (caucasoids), africans (negroids) and ameriesiandians (mongoloids). Most people from the State of Antioquia (4.342.347 inhabitants), belong to the self-designed "paisa" community (Bravo et al., 1996). The paisa community speaks spanish and is geographically located between the central and western branches of the Andean mountains. Anthropological and historical studies describe this population as the most clearly defined in Colombia. Its ethno-historical origin stems most probably from the Spaniards, Jews (christianized sephardim or Marranos), and basques. On the other hand, the admixture with negroid and ameriesiandian populations has been historically documented as low (Bravo et al., 1996). The 95% interval confidence of the estimated negroid racial component included 0%. Thus far, this
community appears to be very interesting for the study of familial aggregation of CLP and for the linkage analysis on extended pedigrees and new mapping strategies, such as MALD, in order to locate major genes involved in susceptibility to CLP.

The present study was undertaken to examine the pattern of inheritance of CLP in a Colombian population. We present evidence that a dominant major locus without multifactorial effects is responsible for the familial aggregation observed in pedigrees ascertained from this community. Some hypotheses on epistatic effects are postulated based in the results of parametric estimation of the t2 parameter when models iterating them were maximized.

METHODS

Sample

Probands affected with CLP were randomly selected according to a sequential sampling strategy among individuals seeking the medical care at the program on CLP in Noel clinic at Medellin, Colombia. Affected individuals were admitted to surgical treatment. Colombian plastic surgeons, orthodontists and geneticists evaluated each one of the probands and a complete clinical and phenotypic record was obtained. Phenotypic classification was defined according to lip and/or palate compromise and unilateral or bilateral affection. Records of facial, maxilla, hands and feet were taken in order to compare these probands with other individuals ascertained world over. Pedigree data and a complete family history of CLP were collected using a specially devised semi-structured questionnaire, which was filled out at a home visit or when interviews were carried out at Noel clinic. Affected relatives pointed out by the probands were clinically analyzed and their phenotypic status was defined.

Ascertainment and complex segregation analysis

The ascertainment probability ($\pi$) was estimated separately from the segregation analysis according to the equation $\sum a(a - 1) / \sum a(r - 1)$ where $a$ is number of probands and $r$ is total number of affected (Simpson, 1983). Complex segregation analysis was carried out according to the unified model of Lalouel et al. (1983), implemented in PONTER computer program (Lalouel and Morton, 1981). The model partitions the total variation in the underlying liability to CLP into three independent components: a diallelic single major locus component, a polygenic background, and a random environmental component. Model parameters are: $q$, the frequency of the high-risk allele $A$; $\theta$, the displacement at the single major locus; $d$, degree of dominance at the major locus, such that $d = 0$ corresponds to a recessive gene, $d = 1$ corresponds to a dominant gene, $0 < d < 1$ corresponds to some degree of additivity and $d = 0.5$ is referred to as codominant; $H$, is the polygenic heritability in the offspring; $Z$, the ratio of adult to childhood heritability; and $t_1$, $t_2$ and $t_3$, the respective probabilities that genotypes $AA$, $Aa$, and $aa$ transmit the allele $A$ (better known as “Eston” probabilities). For example, if the single major locus is mendelian, then $t_1 = 1$, $t_2 = .5$, and $t_3 = 0$, whereas the t’s are equal if there is no transmission of a major effect.

The analysis using pointers only accepts nuclear families as input. Therefore, extended pedigrees were analyzed by dividing them into their component nuclear families. Those nuclear families not containing affected probands but containing affected relatives of the “pointer” (nominal proband) were codified in each sibship considering that the ascertainment probability value $\pi = 1$. Only nuclear families ascertained through pointers with at least one affected individual were included. This last approach was chosen because simulations and empirical results have shown similar results either including or not-including families with unaffected members (Marazita et al., 1992). In addition, as nonsyndromic CLP is a disorder with sex dependent liability, and our preliminary results in the Colombian population have confirmed this assumption (Bravo et al., 1998), two classes of susceptibility were defined, males (0.58 per 500 live births) and females (0.42 per 500 live births). Conditional likelihood was used when maximizing the different models based on
the sex incidences. Neither mortality nor marriage differential risk was taken into account. When the heritability parameter was iterated, different numbers of quadrature points (between five and twenty) were used to reach stable values of likelihood.

### RESULTS

A total number of 95 extended multigenerational pedigrees were analyzed, constituted by 201 nuclear components and 1,136 records. The number of probands was 99 and the total number affected was 234 (144 males and 90 females) (π = 0.42). Family distribution of sibship size was as follows: 1 sib (61 sibships), 2(45), 3(35), 4(21), 5(13), 6(8), 7(3), 8(2), 9(2), 10(5), 11(2), 12(3), 13(1).

Table 1 presents the results of complex segregation analysis of the data. Ten hypothetical models were contrasted using likelihood ratio tests. -2 log likelihood values for each comparison were examined using χ² tests. Parameter estimates corresponding to maximum likelihood models under each set of constraints are shown for each examined model.

The hypothesis of non-familial transmission for CLP in these families (cohort effect) (comparison between models 1 and 9) was rejected (χ²4 df = 929.85, P < 0.0001). The hypothesis of a multifactorial component compared against that of the existence of a major gene only (comparison between model 2 and model 7) was rejected (χ²2 df = 347.97, P < 0.0001). Contrast of the multifactorial hypothesis evaluating the Z parameter (comparison between model 2 and 3) did not show significant differences, meaning that there are not intergenerational differences in the heritability (χ²1 df = 29.43, P < 0.0001). Among the models postulating a major locus (dominant, codominant or recessive), only the recessive model could be rejected (comparison between model 6 and 7) (χ²1 df = 302.95, P < 0.0001). On the contrary, codominant (comparison between models 5 and 7) and dominant models (comparison between models 4 and 7) could not be rejected (χ²1 df = 0.00, P > 0.05 and χ²1 df = 0.39, P > 0.05, respectively). Iteration of the t2 para-

---

**Table 1. Results of complex segregation analysis**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d</td>
</tr>
<tr>
<td>1- No transmission (q = H = 0).</td>
<td>(0)</td>
</tr>
<tr>
<td>Sporadic</td>
<td></td>
</tr>
<tr>
<td><strong>Multifactorial</strong></td>
<td></td>
</tr>
<tr>
<td>2- No cohort effect</td>
<td>(0)</td>
</tr>
<tr>
<td>3- Cohort effect</td>
<td>(0)</td>
</tr>
<tr>
<td><strong>Major locus</strong></td>
<td></td>
</tr>
<tr>
<td>4- Dominant</td>
<td>(1)</td>
</tr>
<tr>
<td>5- Codominant</td>
<td>(0.5)</td>
</tr>
<tr>
<td>6- Recessive</td>
<td>(0)</td>
</tr>
<tr>
<td>7- Unrestricted d</td>
<td>1.0</td>
</tr>
<tr>
<td>8- Unrestricted d and t2</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Mixed model</strong></td>
<td></td>
</tr>
<tr>
<td>9- Unrestricted d</td>
<td>(1)</td>
</tr>
<tr>
<td>10- no transmission of</td>
<td>1.0</td>
</tr>
<tr>
<td>major effect (t's equal)</td>
<td></td>
</tr>
</tbody>
</table>

* d = dominance; t = standard deviations among homozygotes; q = gene frequency; H = Heritability; Z = intergenerational ratio among heritabilities; t1, t2, t3 = Elston probabilities.
meter in the model of major gene yield an abrupt drop in the value of t2 with asymptote in 0. Significant differences were found when models 7 and 8 were compared among them (χ²1 df = 297.69, P > 0.0001). Iteration of the t2 parameter also cause a considerable increment in the frequency of the major gene. Model postulating no polygenic component in the mixed model could no be rejected (comparison of models 7 and 9) (χ²1 df = 0.30, P < 0.05), while the model of no major component in the mixed model was rejected (comparison of models 2 and 9) (χ²3 df = 347.98, P < 0.0001). Finally, the model of non-transmission of major effect (t1 = t2 = t3)(comparison of models 9 and 10) was rejected (χ²1 df = 1003.05, P < 0.0001).

The estimations of penetrance (affection status/genotype) and risk (genotype/affection status) values according to the two susceptibility classes are shown in table 2. It is worth mentioning that the probability to be affected when an individual is heterozygote (those individuals with a greater risk) is greater for males, reaching a value close to 50% (45.23).

**DISCUSSION**

Thus far, the most parsimonious model of inheritance for CLP is that of a major gene (dominant-codominant) without multifactorial effects. From the general unrestricted model (model 9), it is deduced that the gene frequency of this major gene is 0.0012 with a high penetrance. We found that the t2 iteration in the major gene model with unrestricted d, result in a significantly improving of the model likelihood. This aspect can be expressing the interaction of at least two major genes, and make that our analysis will be compatible with that supporting oligogenic interactions. For example, Farral and Holders (1992) found that a monogenic/additive model is strongly rejected. The limited available twin data are also consistent with this model. A “major gene” interacting epistatically with an oligogenic background is shown to be a plausible alternative. In the same way, as has been pointed by Mitchell and Risch (1993), the pattern of recurrence among MZ twins and more remote relatives of CLP probands is not consistent with single major locus inheritance but is compatible with either an MFT model or a model specifying multiple interacting loci. Estimations of penetrance by using conditional probabilities on both the genotype and the status considering the existence of a dominant major gene are presented in table 2. In this table is showed that penetrance for each liability class, males and females is 50% and 30% respectively and that most of individuals being affected by CLP are those individuals being heterozygotes.

Although the results of the present analysis are embedded in the results of many other studies, they could be important in several ways: first, they are compatible with other reports on the mode of inheritance of CLP in other Latin American communities (Chilean population) (Palomino et al., 1997; Blanco et al., 1998b). The hypotheses of ethnic heterogeneity in the CLP risk has been tested in other studies, such as those of Chung et al. (1989), Amidei et al. (1994), Blanco and Rosales (1988), but the results were not conclusive. The results of this analysis could reflect a homogeneous behavior of the major genes which account for the susceptibility to develop CLP in the Latin American admixture populations. Second, the use of linkage analysis for CLP makes the results of this segregation analysis important since an in

<table>
<thead>
<tr>
<th>P(affection/genotype)</th>
<th>Incidence</th>
<th>Threshold</th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>.00120</td>
<td>5.58585</td>
<td>.50030</td>
<td>.50030</td>
<td>.00000</td>
</tr>
<tr>
<td>Females</td>
<td>.00076</td>
<td>6.04485</td>
<td>.31686</td>
<td>.31686</td>
<td>.00000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P(genotype/affection status)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liability = males</td>
</tr>
<tr>
<td>P(G/A)</td>
</tr>
<tr>
<td>P(G/N)</td>
</tr>
<tr>
<td>Liability = females</td>
</tr>
<tr>
<td>P(G/A)</td>
</tr>
<tr>
<td>P(G/N)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Over classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(G/A)</td>
</tr>
<tr>
<td>P(G/N)</td>
</tr>
<tr>
<td>P(G/U)</td>
</tr>
</tbody>
</table>
correctly specified model leads to difficulties. Moreover, misprediction of allele frequencies, especially in the presence of sporadic cases of CLP, will greatly influence the power of any linkage study. Third, empiric risks can be used in clinical practice for counseling. One objective of segregation analysis is to define a more accurate estimate of risk than could be obtained through empiric calculations, which ignore the etiology of the disorder (Houlston et al., 1991). Fourth, as has been pointed out by Lalouel et al. (1983), dominance at the major locus leads to greater correlation between sibs than between parent and offspring. The later observation may, however, result from a variety of other factors such as a common sibling environment, trends of variance components with age, or deviations from assumptions about linearity and additivity effects. More generally, multifactorial transmission may concern environmental as well as genetic effects.

One approach can allow for intergenerational differences in the multifactorial variance components, as well as general transmission through the parameter $r$ (the parent-offspring correlation, conditional on the major genotypes and random residuals). The data can be adjusted so that $V$ is the same in each generation; if $C_A$ and $C_K$ denote variance components due to multifactorial transmission in adults and young, respectively, we may define childhood heritability as $H = C_K/V$, adult heritability as $HZ = C_A/V$, where $Z = C_A/C_K$, and parent-offspring and sib correlations as $rHZ^{1/2}$ and $2r^2H$, respectively. In our case, when the $Z$ parameter was iterated, we observed a significant ($p < 0.0001$) increment of the $-2\ln(L) + C$ (model 3) to expenses of a dramatic fall of $Z$ which can be explained by saying that there are no differences in the multifactorial variance.

REFERENCES


