GENETIC INFERENCES ABOUT PATERNITY IN THE PAISA COMMUNITY FROM ANTIOQUIA, COLOMBIA

INFERENCIAS GENÉTICAS SOBRE ASIGNACIÓN DE PATERNIDAD EN LA COMUNIDAD PAISA DE ANTIOQUIA, COLOMBIA

Maria Luisa Bravo-Aguiar[1] and Mauricio Arcos-Burgos[1,2]

Abstract

This paper compiles gene frequencies of classical, and DNA genetic markers in order to solve the power of non exclusion paternity for the Antioquian community in Colombia. Two ways of quantifying the rarity of non-exclusion of paternity were used. The first way was the Bayesian approach that compute a likelihood ratio of the trio with the putative father as real father versus the trio with the real father as a random male. This likelihood ratio or paternity index can be transformed into a posterior probability if a prior probability of paternity is supplied. The other approach was that of the computation of the probability that a random male would not be excluded by at least one of the tests given the phenotypes of the mother and child. In this case, we carried out the estimation of the paternity index and of the non-exclusion probability. Both of these measures are computed for each locus separately and cumulated over all loci. All analyses are exemplified by using a real example. Finally, we estimated the paternity exclusion probability by using the Ohno *et al.* equation for each locus with n alleles.

Key words: paternity testing, genetic markers, Colombia, genetics, Antioquia, population genetics, blood groups, polymarker.

Resumen

Este artículo compila información sobre la distribución de las frecuencias genéticas de marcadores genéticos clásicos y de DNA con la finalidad de observar el comportamiento del parámetro de no-exclusión en disputas de paternidad originadas en individuos que pertenecen al departamento de Antioquia, Colombia. Primero usamos la aproximación Bayesiana, que computa la probabilidad del trío en problema tomando al padre real como el padre putativo, comparada contra la probabilidad de un trío con un padre aleatorio tomado de la población problema. Esta razón de verosimilitud o índice de paternidad puede ser transformada en una probabilidad posterior si una probabilidad a priori de paternidad es entregada. La otra aproximación que usamos fue el cómputo de la probabilidad de que un hombre aleatorio no sea excluido por al menos una de las pruebas biológicas realizadas teniendo como condición los fenotipos de la madre y el niño. En este caso, estimamos el índice de paternidad y la probabilidad de no-exclusión. Todos los parámetros fueron estimados para cada locus y acumulados sobre todos los loci. Se usó un ejemplo real como aproximación práctica. Finalmente estimamos el poder de exclusión de la paternidad de acuerdo con la ecuación de Ohno *et al.* para cada locus con n alelos.

Palabras clave: pruebas de paternidad, marcadores genéticos, Colombia, genética, Antioquia, genética de poblaciones, grupos sanguíneos.

INTRODUCTION

In paternity disputes, it is necessary to determine whether a particular man is the father of a particular child. Classical considerations of such questions were limited to excluding a man from paternity of a

Recibido: febrero de 1999; aprobado para publicación: agosto de 1999.

[1] Forensic Sciences Laboratory, Biology Department, University of Antioquia, Medellin, Colombia.

[2] Population Genetics and Genetic Epidemiology Group, Biology Department, University of Antioquia, Apartado 1226, Medellin, Colombia. E-mail: omarcos@matematicas.udea.edu.co.

child that has neither of the alleles of man at some locus (Weir, 1990). On the other hand, when the particular man has some of the alleles carried by the child, the probability of not exclusion will be a function of the estimated allelic frequencies of the population where the individuals were ascertained. In the same way, when several analyzed loci have been compatible, the non-exclusion probability will be the disjunctive function of allelic frequencies for each locus that has been involved in the probe. Must be apparent that the higher either polymorphism or heterozygosity, the higher the power to exclude an individual as putative father does.

In most of the cases submitted to the forensic analysis, we want obtain the lower probability that an accused individual being not guilty in nature cannot be discriminated from another individual belonging to the same population. This probability can be diminished as low as either the number of loci or the power of each one of locus will be incremented. The increasing availability of DNA diagnostic loci, anywhere minisatellites, microsatellites, or STR's loci, has improved the performance in discriminating individuals from the remaining population as function of the non-exclusion probability.

In our case, although the population of Colombia stemmed mostly from the admixture of Europeans (Caucasoid), Africans (Negroid) and Amerindians (Mongoloid) (Agudelo, 1986), most population belonging to Antioquia, a department of Colombia, whose capital is Medellin belongs to the self-designed "paisa" community (Agudelo, 1986). The paisa community is geographically located within the area outlined by the Central and Western branches of the Andean mountains encompassing the actual states of Antioquia, Caldas, Risaralda, Quindío, and the northern lands belonging to the states of Valle and Tolima. This community has based their economy in the agriculture and today represent one of the most important markets of the country. Anthropological and historical studies describe this population as the most clearly defined in Colombia. They are predominantly catholic, very endogamic and conservative. Their familial structure is matriarchal in nature and their offspring is characterized by a considerable big sib-ships. 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 Amerindian populations has been historically documented as low (Agudelo, 1986; Parsons, 1949; Bravo *et al.*, 1996).

Here we compiles gene frequencies of classical, and DNA genetic markers in order to solve the power of non exclusion paternity, for each locus and the joined set of markers when them are used in paternity disputes.

MATERIALS AND METHODS

Samples and genotyping. Several sets of data were used in order to estimate the gene frequencies for the different genetic markers. In summary: four data sets were used to perform the analysis.

In a first set, blood samples were collected in bulbs without anticoagulant by vein puncture and tested within 48h of sampling. This sample belongs to 449 unrelated patients (225 males and 224 females). They came to the Forensic Sciences Laboratory of the University of Antioquia at Medellin. All of them were born at Antioquia department, and were among 15 to 30 years old. Their parents and grandparents were from Antioquia. Plasma was isolated for conventional centrifugation and the plasmatic proteins transferrin (TF), haptoglobin (HPA*), specific group component (GC, DBP) were separated for electrophoresis in polyacrilamid gels according to standard techniques (Hames, 1987). In this set, erythrocyte groups Duffy (FY) and P system (P) were typed according with the micro method described by Sussman (1976). Ortho-Pharmaceutical Corporation, New Jersey, USA, provided the antisera.

In a second set of 500 individuals (250 males and 250 females), with similar features, and ascertained at the same place as the first set, the ABO, RH, MNS, Kidd (JK) and Kell (K) erythrocyte groups were typed. The conditions and methods for genotyping are the same that were described above. For MNS, JK and K, only 251, 80 and 80 individuals respectively were typified according to antisera availability.

In a third set, 80 unrelated individuals assisting to the Genetic Laboratory of Forensic Sciences were typified to determine polymorphism at DNA markers namely LDLR, GYPA, HBGG, D7S8, GC and DQα.

Maximum likelihood estimates of gene frequencies of protein plasmatic markers and erythrocyte markers were obtained using the programs MAXLIK and MENDEL.

Paternity testing. There are two ways of quantifying the rarity of this event. The first way is to take a Bayesian approach and compute a likelihood ratio of the trio with the putative father as real father versus the trio with the real father as a random male. This likelihood ratio or paternity index can be transformed into a posterior probability if a prior probability of paternity is supplied. Some statistical geneticists refuse to take this Bayesian step and simply insist that the paternity index has a fundamental importance of its own. Another approach to the problem is to compute the probability that a random male would not be excluded by at least one of the tests given the phenotypes of the mother and child. This may not be as powerful, but it is free from controversy. In this case, we carried out the estimation of the paternity index and of the non-exclusion probability. Both of these measures are computed for each locus separately and cumulated over all loci. All analyses are exemplified by using a real example.

Finally, we estimated the paternity exclusion probability by using the Ohno *et al.* equation (1996) for each locus with n alleles. If Pr(PE)=(Paternity Exclusion Probability), then

$$\Pr(PE) = \sum_{i=1}^{n} pi(1-pi)^2(1-pi+pi^2) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} pipj(pi+pj)(1-pi-pj)^2 \qquad (2.1)$$

Equation 2.1 permits us to observe the power of each locus in discriminating genotypes belonging from a population of individuals. Appendix 1 shows its implementation in a Fortran 77 program.

RESULTS

In the table 1 the gene frequencies distributions for all set of genetic markers to the population from where the sample was drawn is showed. It is worth to mention that neither the effects of linkage disequilibrium nor the effects of population subdivision were considered to determine the final estimates.

The table 2 shows the genotypes at all loci to the trio (child, mother and the putative father) submitted to analysis.

The table 3 shows the respective values to the index of paternity and the value to the non-exclusion probability for each locus and their respective accumulated values.

By reporting these non-exclusion probabilities for each locus separately, you can see those loci that are critically important in the paternity evaluation. The cumulative measures, however, are the ones that should be quoted.

In table 4 are presented the probability values of paternity exclusion by using the Ohno *et al.* equation. This estimated value can be very useful in determinating empirically or *a priori* those genetical markers that with higher power to discriminate individuals.

DISCUSSION

It has been showed that the higher the population genetic diversity the higher the power to discriminate individuals. The Antioquian population shows an extraordinary homogeneous genetic structure making the discrimination of individuals a hard enterprise. In this article is showed that scarcely, although the extraordinary number of loci with high polimorphism and with a high power, many times, as is exemplified, levels closed to 0.03 are obtained. These results indicates that, necessarily more loci must be typed, being preferential those as microsatellite loci. In the next paper, we will show the improvement of this paternity inference by using STR's.

Table 1. Allele frequencies in samples from the Antioquia population (paisas)

LOC	HLA-A	HLA-B		HLA-C		BF		RH		MNS		ABO		DQ	
N	87	87		87		87		500		251		500		80	
AL		AL		AL		AL		AL		AL	AL			AL	
1	.109	5	.149	1	.086	S	.632	CDE	.010	MS	.273	A	.190	A	0.134
2	.172	7	.063	2	.023	F	.195	CDe	.396	Ms	.289	B	.042	B	0.164
3	.149	13	.040	3	.172	S*	.023	cDE	.166	NS	.122	O	.768	C	0.084
10	.063	14	.040	4	.259	F1	.063	cDe	.060	Ns	.317			D	0.111
11	.109	16	.040	5	.086	F*	.086	CdE	.001					E	0.269
19	.086	18	.132	6	.023			Cde	.021					F	0.234
24	.086	27	.023	7	.040			cdE	.014					LDLR	
28	.149	35	.132	X	.305			cde	.332					A	0.3520
X	.063	37	.023											B	0.6480
		40	.109											GYPA	
		44	.086											A	0.4260
		49	.023											B	0.5740
		56	.023											HBGG	
		57	.023											A	0.3700
		X	.040											B	0.5740
														C	0.0560

LOC	FY	P		HPA*		T		GC-DBPK		K		JK			
N	449	449		449		449		449		80		80		D7S8	
AL		AL		AL		AL		AL		AL		AL		A	0.5190
A	.539	P1	.550	1	.455	C	.994	1	.669	K	.031	a	.519	B	0.4810
B	.460	P2+p	.450	2	.545	B	.001	2	.331	k	.969	b	.481	GCC	
(-)	.001					D	.004							A	0.1300
														B	0.1850
														C	0.6850

LOC = locus; AL = allele; N = sample size; S* = S0.7 (BF)

Table 2. A practical example father (acussed), mother and son with their respective genotypes for each one of the used markers

2. Mother	O k P1 A A BD	DccEe A AB A AB	MMSs AB C A BC
3. Child	O k P1 A A BD	dccee A AB AB B	MMSs AB C A BC
4. Accused	O k P1 A AB BD	DCcee A AB B AB	MNss AB C AB BC
Format as markers has been disposed for each individual	ABO KELL P GC HBGG DQ	RH KIDD HP LDLR HBGG	MNS DUFFY TF GYPA GCC

Table 3. Results of estimations for each one of the paternity parameters analyzed

	PATINDEX	CUMINDEX	PAT PROB	CUM PROB
ABO	.1335D+01	.1335D+01	.9370D+00	.9370D+00
RH	.1265D+01	.1689D+01	.5538D+00	.5189D+00
MNS	.8913D+00	.1505D+01	.8073D+00	.4189D+00
KELL	.1032D+01	.1554D+01	.9990D+00	.4185D+00
KIDD	.1927D+01	.2993D+01	.7686D+00	.3217D+00
DUFFY	.1001D+01	.2996D+01	.1000D+01	.3217D+00
P	.1050D+01	.3147D+01	.1000D+01	.3217D+00
HP	.1000D+01	.3147D+01	.1000D+01	.3217D+00
TF	.1006D+01	.3166D+01	.1000D+01	.3217D+00
GC	.1495D+01	.4733D+01	.8904D+00	.2864D+00
LDLR	.1543D+01	.7304D+01	.8761D+00	.2509D+00
GYPA	.1174D+01	.8573D+01	.6705D+00	.1683D+00
HBGG	.1351D+01	.1158D+02	.6031D+00	.1015D+00
D7S8	.1040D+01	.1204D+02	.7306D+00	.7414D-01
GCC	.1495D+01	.1800D+02	.8904D+00	.6602D-01
DQ	.3620D+01	.6517D+02	.4762D+00	.3143D-01

Tabla 4. Probability of excluding paternity with any number of codominant alleles according with the Ohno *et al.* equation (1996)

HLA-A	0.7696	LDLR	0.1761
HLA-B	0.7426	GYPA	0.1841
HLA-C	0.0563	HBGG	0.2594
BF	0.3200	D7S8	0.1873
MNS	0.6388	GCC	0.2618

Apendix I

```
c     program EXPATER//FORTRAN 77
c     ---------------------------------------------------------
c     This program computes the paternity exclusion probability
c     by using the Ohno et al. equation (1982)
c     for codominant loci with n alleles.
c     Implemented by Mauricio Arcos-Burgos
c     ---------------------------------------------------------
c     Declaration of the variables:
      real*8 p(1,30), s1, s2, s3
      integer*4 n, m
c     ---------------------------------------------------------
c     Opening the input and output files:
      open(1,file='expater.inp',status='old')
      open(2,file='expater.out',status='new')
c     ---------------------------------------------------------
c     Read the alleles (n)
      read (1,*)n
      m=n-1
      do 5 j=1,n
        read (1,*) p(1,j)
5     continue
c     Writing the headers for the output file:
      write (2, 30)
30     format (2x,'****Paternity Exclusion****:')
      write (2, 40)
40     format (5x,'The Alleles are:')
      write (2, 50) (p(1,i), i = 1,n)
50     format(10f10.4)
c     Computation of sums
      s1=0.0
      do 150 i = 1,n
        s1=s1+s2+p(1,i)*(1-p(1,i))**2*(1-p(1,i)+p(1,i)**2)
        do 170 k = 1,m
          s2=0.0
          s2=s2+s3
          j=i+1
            s3=0.0
          do 180 j = 1,n
            s3=s3+p(1,k)*p(1,j)*(p(1,k)+p(1,j))*
     1      (1-p(1,k)-p(1,j))**2
180       continue
170     continue
150   continue
      PE=s1
      write (2, 191)
191    format (2x,'****Paternity Exclusion****:')
      write (2,192)PE
192    format (20X,1f8.4)
      stop
      end
```

REFERENCES

Agudelo LE. 1986. *Génesis del pueblo antioqueño.* Era Cósmica Ediciones. Santafé de Bogotá, DE. Colombia.

Bravo ML, Valenzuela YC, Arcos-Burgos M. 1996. Polimorphisms and phyletic relationships of the paisa community from Antioquia (Colombia). *Gene Geography* 10:11-17.

Hames BD. 1987. *Gel electrophoresis of proteins.* IRL Press, Oxford, Washington, DC.

Ohno Y, Sebetan IM, Akaishi S. 1996. A simple method for calculating the probability of excluding paternity with any number of cdominant alleles. *Forensic Sci Internat* 19:93-98.

Parsons JJ. 1949. *Antioqueño colonization in western Colombia.* University of California Press. Berkeley.

Sussman LN. 1976. *Paternity testing by blood grouping.* C. Thomas Publisher, USA.

Weir B. 1990. Genetic data analysis. Methods for discrete population genetic data. Sinauer Associates, Inc. Publishers. Sunderland, Masachusetts, 185 p.