

# COMPUTACION DE COEFICIENTES DE CONSANGUINIDAD Y DE PARENTESCO

Computer calculation of inbreeding and relationship coefficients

Calcul avec l'ordinateur des coefficients de consanguinité et de parenté

P. ALFONSO PONCE \*

## INTRODUCCIÓN

El estudio y medida de la consanguinidad tiene un notable interés en cualquier plan de mejora genética o evolución artificial porque un alto nivel de consanguinidad es necesario si se quieren mantener individuos estrechamente relacionados con un antecesor deseable; ayuda a descubrir genes recesivos perjudiciales para que puedan ser expurgados de una población; forma islotes familiares homogéneos y distintos entre sí en el seno de una población, semejantes a los producidos por los aislamientos geográficos, posibilitando una selección interfamiliar más efectiva que si no hubiera consanguinidad. Por otra parte, la medida de la consanguinidad tiene también gran importancia para controlar la inseminación artificial ganadera, técnica cada día más extendida que al permitir una gran disminución del número de machos reproductores está conduciendo fatalmente, en mayor o menor plazo, al aumento de parentescos masales y promedios de consanguinidad en las poblaciones que se aplica, con los consiguientes efectos anómalos por concentración de genes desfavorables y con efectos depresivos como disminución de fertilidad y viabilidad.

S. WRIGHT (1922) es el primero en medir la consanguinidad generalizando los resultados de R. A. FISHER (1918) e introduciendo los coeficientes de consanguinidad y parentesco. Define al coeficiente de consanguinidad como la correlación lineal entre los dos gametos que al unirse forman un cigoto consanguíneo, y al coeficiente de parentesco como la correlación lineal entre dos cigotos no consanguíneos. Para el cálculo sobre genealogías utiliza las nociones de coeficiente viario y camino de parentesco. G. MALÉCOT (1948) hace un planteamiento radicalmente

---

\* Departamento de Genética y Mejora, Facultad de Veterinaria, Universidad Complutense de Madrid, Ciudad Universitaria, Madrid-3, España.

distinto de S. WRIGHT y seguidores. Considera al coeficiente de consanguinidad no como una correlación lineal, sino como la probabilidad de un suceso, es decir, la probabilidad de que dos genes homólogos, tomados al azar en dos zigotos, sean idénticos por descendencia mendeliana, en ausencia de mutación, de un mismo gen antepasado, y al coeficiente de parentesco como la probabilidad de que dos genes homólogos, tomados al azar en dos zigotos, sean idénticos por descendencia mendeliana, en ausencia de mutación, de un mismo gen antepasado. Para el cálculo sobre genealogías utiliza la noción de cadena de parentesco. M. GILLOIS (1964) generaliza los trabajos de G. MALÉCOT y muestra que el conocimiento del conjunto de relaciones genéticas entre dos individuos se puede resumir por medio de 15 parámetros que denomina coeficientes de identidad que son las probabilidades correspondientes a las 15 situaciones posibles de identidad entre los cuatro genes homólogos de dos zigotos. Para determinar cada situación de identidad utiliza la noción de caso de ascendencia que es el conjunto de caminos de procedencia u orígenes de los cuatro genes homólogos. Considera la consanguinidad y el parentesco como la traducción exclusiva del estado de relaciones de ascendencia; y la identidad, como la descripción de estados bioquímicos de genes que pueden ser o no ser copia de un mismo modelo.

En el cálculo de coeficientes de parentesco y consanguinidad por los métodos más importantes es necesario encontrar los distintos caminos, cadenas de parentesco, casos de ascendencia, lo que en genealogías extensas y complejas resulta muy laborioso y se presta al error frecuente de tomar por válidos los repetidos. Con relación a este problema nos ocuparemos aquí de las genealogías de individuos emparentados y consanguíneos consideradas como redes matemáticas en las que existe la estructura *arborescencia genealógica* (P. ALFONSO PONCE, 1971) de propiedades características para poderla detectar sin error posible en genealogías por extensas y complejas que sean, lo que permite un método de cálculo seguro de coeficientes de parentesco y consanguinidad en un ordenador electrónico.

### 1. *La arborescencia genealógica*

Toda genealogía es una red o grafo,  $G = (X, \Gamma x)$ , porque se compone de un conjunto  $X$  de individuos y de una aplicación  $\Gamma x$  de  $X$  en  $X$  que es el conjunto de arcos o relaciones binarias de ascendencia  $\mathcal{A}$  entre ellos.

El conjunto de antepasados y relaciones de ascendencia comunes a dos individuos emparentados,  $i$  y  $j$ , es la intersección de sus redes genealógicas,  $G_i \cap G_j$ , donde  $G_i = (A, \Gamma_A)$  y  $G_j = (B, \Gamma_B)$ . Por la intersección de dos redes genealógicas,  $G_i \cap G_j$ , resulta una red  $G_{ij}$  (fig. 1).

Desde la intersección  $G_i \cap G_j$  hasta la base de la red  $G_i$  y la base de la red  $G_j$  se pueden analizar los posibles sucesos de transmisión de genes idénticos, los cuales ocurren solamente a través de determinados sectores de la red  $G_{ij}$ , es decir, no todos los individuos pertenecientes a la intersección  $G_i \cap G_j$  tienen interés para el cálculo del parentesco entre  $i$  y  $j$  y consecuentemente para la consanguinidad de un hijo de éstos, sino exclusivamente aquellos que, extraídos de la red  $G_{ij}$ , junto con sus descendientes, hasta la base de la red  $G_i$  y la base de la red  $G_j$ , formen

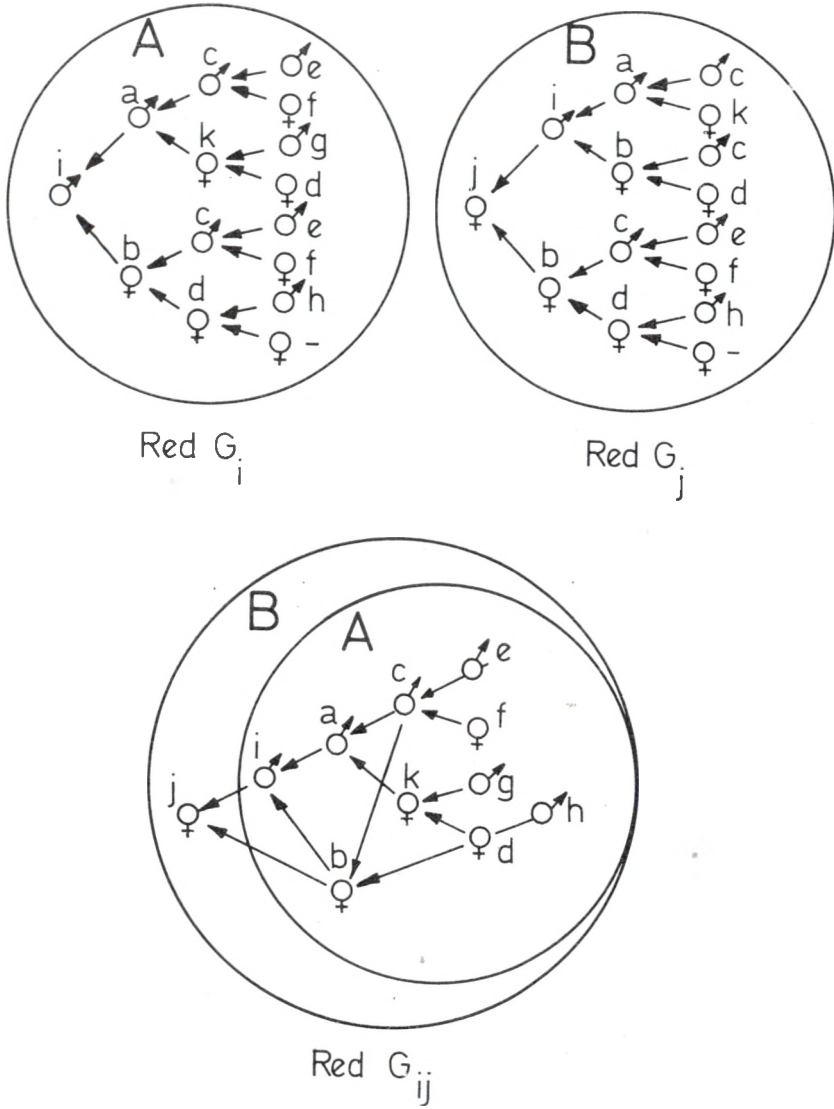


FIG. 1: Intersección de redes genealógicas emparentadas

lo que denominamos una arborescencia genealógica\*. Consiste en una subred de  $G_{ij}$ . Está formada por una raíz o antepasado común  $x_i$  y por dos caminos  $\mu_1$  y  $\mu_2$  que son la secuencia de arcos o relaciones binarias de ascendencia,  $\mathcal{A}$ , entre los des-

\* Adoptamos este término por la gran analogía con lo que en teoría de redes se conoce por arborescencia.

endientes de  $x_i$  hasta la base de la red  $G_i$  y la base de la red  $G_j$ . Tiene cuatro propiedades algebraicas características:

- 1.ª todo nudo distinto de  $x_i$  es la extremidad inicial de un solo arco;
- 2.ª todo nudo distinto de  $x_i$  es la extremidad terminal de un solo arco;
- 3.ª  $x_i$  no es la extremidad terminal de ningún arco de la subred, y
- 4.ª los caminos  $\mu_1$  y  $\mu_2$  no tienen ningún arco en común.

## 2. Método de cálculo de coeficientes de parentesco y consanguinidad por medio de arborescencias genealógicas.

Las cuatro propiedades de la arborescencia genealógica nos han sugerido un método de cálculo de coeficientes de parentesco y consanguinidad cuyos pasos fundamentales, desarrollados sobre el ejemplo de la figura 1 (genealogías de los padres del históricamente famoso toro *Comet*), son los siguientes:

1.º Disponiendo en dos ejes de coordenadas (fig. 2) los dos conjuntos de ascendencia  $A$  y  $B$  por el siguiente orden lineal desde el origen: 1 (individuo), 2 (padre), 3 (madre), 4 (abuelo paterno), 5 (abuela paterna), 6 (abuelo materno), 7 (abuela materna), etc., se puede realizar el producto cartesiano de conjuntos  $A \times B$  y por tanto todas las comparaciones posibles entre antepasados. Trazando, en cada relación binaria de identidad, desde la abscisa y ordenada, dos cotas perpendiculares se puede situar un antepasado común a las dos genealogías, representado por un punto físico en el plano e identificado por un par ordenado  $(x, y)$ .

2.º Cada antepasado común o gráfica de par ordenado  $(x, y)$  se extrae junto con todos sus descendientes en  $A$  y  $B$  para comprobar si la subred extraída cumple las cuatro propiedades de la arborescencia genealógica. Tal extracción es sumamente sencilla; para ello basta con examinar el orden numérico lineal de una y otra genealogía en la abscisa y ordenada, y en seguida se ve que los descendientes del par ordenado  $(x, y)$  son:

$$(x, y) \begin{cases} \nearrow (x/2, 0) \rightarrow (x/4, 0) \rightarrow (x/8, 0) \\ \searrow (0, y/2) \rightarrow (0, y/4) \rightarrow (0, y/8) \end{cases}$$

despreciando la parte decimal del resultado de las divisiones si  $x$  e  $y$  son números impares. Por ejemplo, los descendientes del par ordenado  $(11, 7)$ , que en la figura 2 corresponde al antepasado común  $d$ , son:

$$(11, 7) = d \begin{cases} \nearrow (5, 0) = k \rightarrow (2, 0) = a \rightarrow (1, 0) = i \\ \searrow (0, 3) = b \rightarrow (0, 1) = j \end{cases}$$

En la parte inferior de la figura 2 se muestran todas las subredes de la red  $G_{ij}$  extraídas por este procedimiento, de entre las que es muy fácil reconocer por sus propiedades a las arborescencias genealógicas:  $A_1$ ,  $A_2$ ,  $A_3$  y  $A_4$ .

3.º Se calculan los valores de probabilidad de cada arborescencia encontrada, teniendo en cuenta que la probabilidad de cada arco es igual a  $1/2$ . Por tanto, el coeficiente de parentesco entre  $i$  y  $j$ ,  $R_{ij}$ , siendo  $i$  consanguíneo con un valor  $F_i = 0,1875$ , es:

$$\begin{aligned} R_{ij} &= Pr(A_1) (1 + F_i) + Pr(A_2) + Pr(A_3) + Pr(A_4) = \\ &= 1/2 (1 + 0,1875) + (1/2)^2 + (1/2)^4 + (1/2)^5 = 0,93750. \end{aligned}$$

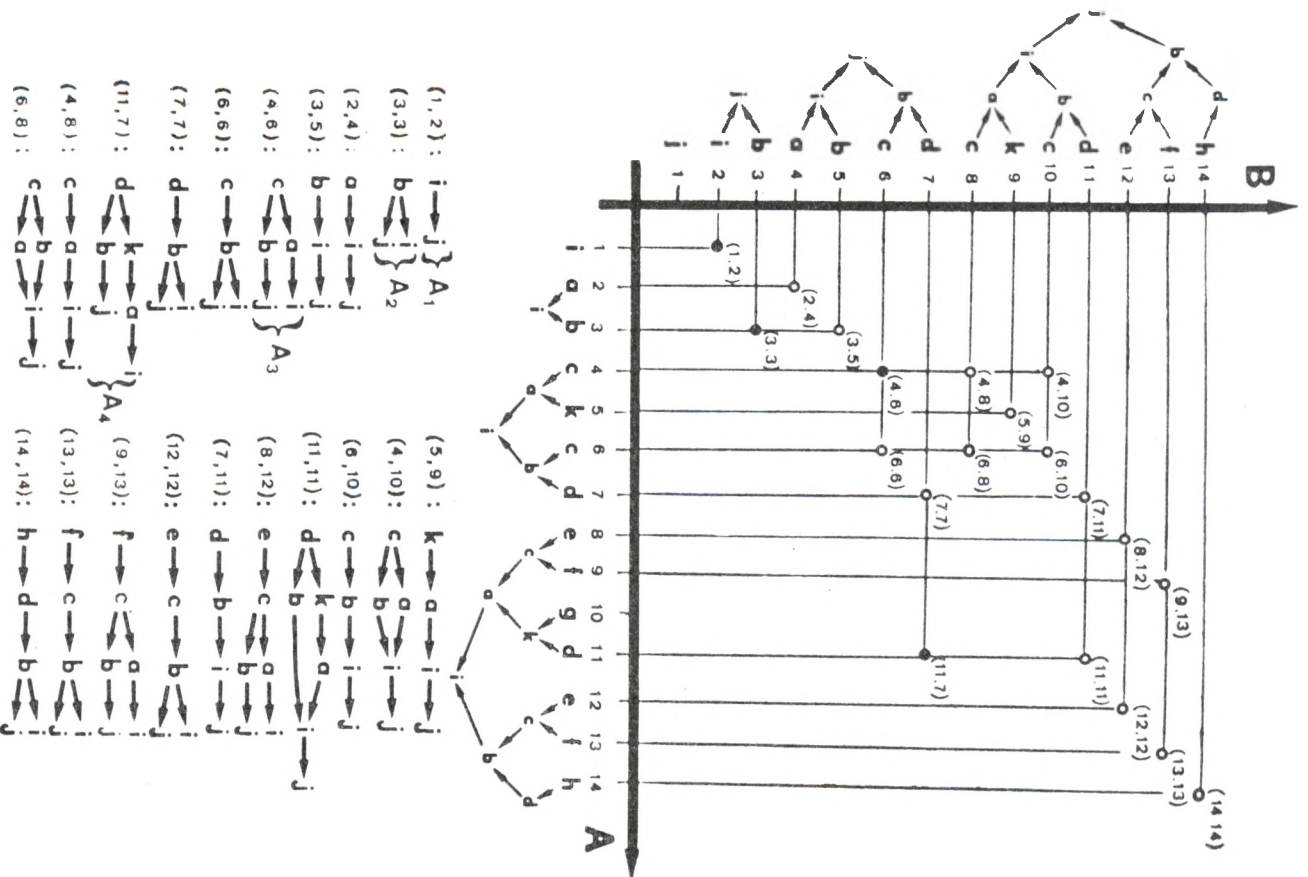


FIG. 2

y el coeficiente de consanguinidad del hijo de  $i$  y  $j$  (toro Comet) es:  $F = 1/2 (0,93750) = 0,46875$ .

### 3. Computación electrónica. Programa Parcon

El programa Parcon, escrito en lenguaje FORTRAN IV, tiene por objeto calcular, por el método anterior, coeficientes de parentesco hasta la 4.<sup>a</sup>, 5.<sup>a</sup> o 6.<sup>a</sup> generación y coeficientes de consanguinidad hasta la 4.<sup>a</sup>, 5.<sup>a</sup>, 6.<sup>a</sup> o 7.<sup>a</sup> generación en un ordenador IBM 7090 o en un IBM 1130; también existe una versión en FORTRAN II-D para el IBM 1620.

*Entrada de datos:* Los números de identificación de los individuos de una genealogía (fig. 3) se perforan todos con cuatro dígitos decimales en fichas estándar IBM de 80 columnas y 12 filas de la siguiente forma:

Columnas	Contenido
1	Números de generaciones de la genealogía
2	En blanco
3 - 6	$X$ (individuo)
7	En blanco
8 - 15	$N_i$ y $M_i$ (padre y madre)
16	En blanco
17 - 32	$N_2, N_3, M_2$ y $M_3$ (abuelos paternos y maternos)
33	En blanco
34 - 65	$N_4, N_5, N_6, N_7, M_4, M_5, M_6$ y $M_7$ (bisabuelos paternos y maternos)
66 - 80	En blanco

Las columnas en blanco separan generaciones. Todos los números de identificación que se desconocen se perforan con ceros.

Si la genealogía llega hasta la 5.<sup>a</sup> generación se perfora una segunda ficha:

Columnas	Contenido
1 - 64	$N_8, N_9, \dots, N_{14}, N_{15}$ y $M_8, M_9, \dots, M_{14}, M_{15}$ (tatarabuelos paternos y maternos)
65 - 80	En blanco

Si la genealogía llega hasta la 6.<sup>a</sup> generación, se perforan en una tercera y cuarta ficha los retatarabuelos paternos y maternos, respectivamente:

Columnas	Contenido
1 - 64 65 - 80	$M_{16}, M_{17}, \dots, M_{30}, M_{31}$ En blanco
1 - 64 65 - 80	$N_{16}, N_{17}, \dots, N_{30}, N_{31}$ En blanco

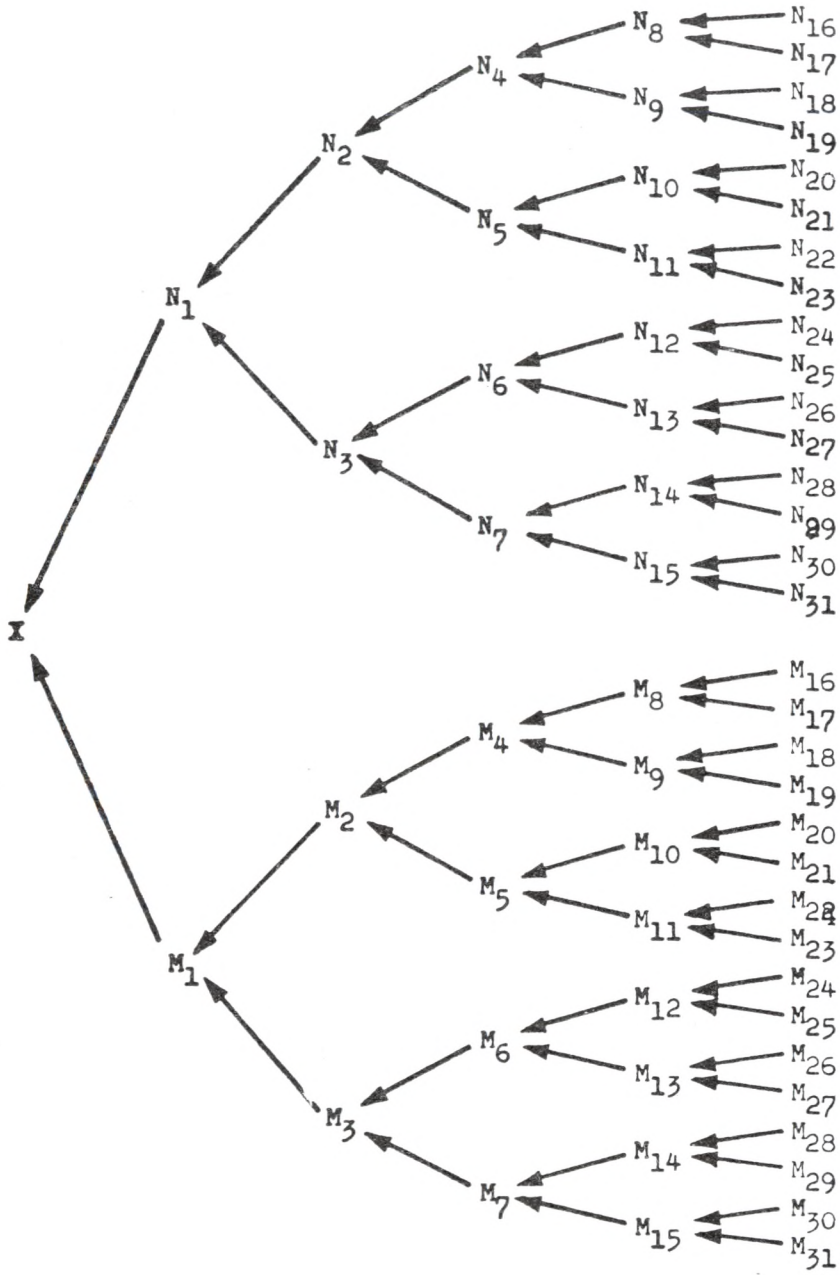


FIG. 3

También las genealogías se pueden componer y perforar automáticamente en el ordenador electrónico partiendo de un fichero de tripletes de individuos de una familia, de una población, en el que cada triplete sean los números de identificación del individuo, de su padre y de su madre.

En cada genealogía se perforan, además, dos fichas para aquellos antepasados que sean consanguíneos con el fin de poder calcular la probabilidad adicional a causa de ello. En la primera, los números de identificación de los antepasados que son consanguíneos; y en la segunda, sus correspondientes valores de consanguinidad. Todos los números son de cuatro dígitos decimales.

Delante de un paquete de fichas genealógicas se coloca una ficha con un número de control perforado en la columna 1. Si tal número es 2, el ordenador calculará coeficientes de parentesco en cada pareja de genealogías que le sigan; si el número es 1, calculará coeficientes de consanguinidad sobre cada una de las genealogías que le sucedan.

La última ficha del paquete de datos es una ficha de control, en blanco, para que finalice el proceso cíclico del programa.

*Descripción del programa:* El ordenador puede leer sobre un mismo paquete de datos, las genealogías una a una, o dos a dos, según se quiera, respectivamente, calcular coeficientes de parentesco o de consanguinidad. El proceso del programa Parcon sigue los pasos 1.º, 2.º y 3.º del método descrito anteriormente, teniendo en cuenta lo siguiente: *a)* según el número de generaciones que tenga cada genealogía, el ordenador ajusta el número de antepasados a comparar, al objeto de reducir tal operación a lo estrictamente necesario; *b)* también con el fin de economizar tiempo de cálculo, realiza todas las comparaciones entre antepasados del mismo sexo; *c)* la comparación entre individuos se hace por sustracción de sus números de identificación; si el resultado es cero, se suman para verificar si se trata de un antepasado común o de dos antepasados desconocidos. A medida que el ordenador finaliza el cálculo de un caso de parentesco o de consanguinidad, imprime su identificación y resultado y se prepara para el siguiente, que puede ser de parentesco o consanguinidad.

## RESUMEN

En las genealogías emparentadas, consideradas como redes matemáticas, existe la estructura arborescencia genealógica con las suficientes propiedades genéticas y matemáticas características para poderla detectar sin error posible en genealogías por extensas y complejas que sean. Con ello se facilita el análisis probabilístico de sucesos de transmisión de genes idénticos y se establece un nuevo método seguro de cálculo de coeficientes de parentesco y consanguinidad en un computador electrónico IBM 7090.

## SUMMARY

In the related genealogies, considered as mathematical networks, an arborescent genealogical structure exists with sufficient genetic properties and mathematical characteristics to be able to detect it without any possible error in genealogies however extensive and complex they may be. With it is facilitated the probabilistic



analysis of the transmission lapses of identical genes and a new reliable method of calculation of coefficients of relationship and consanguinity is established on the electronic computer IBM 7090.

### RESUME

Chez les généalogies emparentées, considérées comme des réseaux mathématiques, il existe la structure arborescente généalogique avec les propriétés génétiques et mathématiques suffisantes pour la pouvoir détecter sans erreur possible, pour extenses et complexes que les généalogies mêmes soient elles. De cette façon est facilité l'analyse de probabilité des événements de transmission de gènes identiques, en établissant une nouvelle et sûre méthode de calcul des coefficients de parenté et de consanguinité dans un ordinateur électronique IBM 7090.

### BIBLIOGRAFIA

- ALFONSO PONCE, P. (1971): Redes genealógicas y computación de coeficientes de parentesco y consanguinidad. *Arch. Zool.*, 20, 79, 257-299.
- BRIONES, F. (1969): *Fortran IV*. Centro de Cálculo de la Universidad de Madrid.
- FISHER, R. A. (1918): The correlation between relatives on the supposition of mendelian inheritance. *Trans. Roy. Soc. Edinburgh*, 52, 399-433.
- GILLOIS, M. (1964): *La relation d'identité en génétique*. Thèse Fac. Sciences. Paris.
- JACQUARD, A. (1966): Logique du calcul des coefficients d'identité entre deux individus. *Population*, 4, 751-776.
- LI, C. C. (1956): The concept of path coefficient and its impact on population genetics. *Biometrics*, 12, 190-210.
- MALÉCOT, G. (1948): *Les Mathématiques de l'Hérédité*. Masson et Cie., Paris.
- WRIGHT, S. (1922): Coefficients of inbreeding and relationship. *Amer. Nat.*, 56, 330-338.
- WRIGHT, S. (1934): The method of path coefficients. *Ann. Math. Statist.*, 5, 161-215.

# LISTADO DEL PROGRAMA PARCON EN FORTRAN IV

PARCON - EFN SOURCE STATEMENT - IFN(S) -

06

```

C PROGRAMA PARCON PARA DETERMINACION DE COEFICIENTES DE PARENTESCO Y
C CONSANGUINIDAD HASTA LA CUARTA, QUINTA O SEXTA GENERACION
  DIMENSION N(63),M(63),NAC(28),CA(28)
  45 FORMAT(32X,63HINDIVIDUOS  NUMERO DE GENERACIONES  COEFICIENTE DE
    1 PARENTESCO//)
  46 FORMAT(32X,67HINDIVIDUO  NUMERO DE GENERACIONES  COEFICIENTE DE
    1 CONSANGUINIDAD//)
  55 FORMAT (11)
  49 FORMAT (11,1X,14,1X,2I4,1X,4I4,1X,8I4)
  50 FORMAT (16I4)
  51 FORMAT (14(1X,14))
  52 FORMAT (14F5.4)
  53 FORMAT (32X,14,2X,14,13X,11,22X,F7.5)
  54 FORMAT (35X,14,16X,11,22X,F7.5)
    1 READ(5,55)NC
      IF(NC-1)58,59,60
C ESCRIBE CABECERA DE COEFICIENTES DE PARENTESCO
  60 WRITE(6,45)
C LEE GENEALOGIA DE N(1) NUMERO DE GENERACIONES, PADRES, ABUELOS Y
C BISABUELOS
  2 READ(5,49)NG,(N(I),I=1,15)
    IF(NG)1,1,3
  3 IF(NG-4)9,9,70
C LEE TATARABUELOS DE N(1)
  70 READ(5,50)(N(I),I=16,31)
    IF(NG-5)9,9,71
C LEE RETATARABUELOS DE N(1)
  71 READ(5,50)(N(I),I=32,47)
    READ(5,50)(N(I),I=48,63)
C LEE ANTEPASADOS CONSANGUINEOS DE N(1) Y SUS COEFICIENTES DE CONSANGUI-
C NIDAD
  9 READ(5,51)(NAC(K),K=15,28)
    READ(5,52)(CA(K),K=15,28)
C LEE GENEALOGIA DE M(1) NUMERO DE GENERACIONES, PADRES, ABUELOS Y
C BISABUELOS
  72 READ(5,49)NG,(M(I),I=1,15)
    IF(NG-4)100,100,73
C LEE TATARABUELOS DE M(1)
  73 READ(5,50)(M(I),I=16,31)
    IF(NG-5)100,100,74
C LEE RETATARABUELOS DE M(1)
  74 READ(5,50)(M(I),I=32,47)
    READ(5,50)(M(I),I=48,63)
    GO TO 100
C ESCRIBE CABECERA DE COEFICIENTES DE CONSANGUINIDAD
  59 WRITE(6,46)
C LEE GENEALOGIA DE IX NUMERO DE GENERACIONES, PADRES, ABUELOS Y
C BISABUELOS
  6 READ(5,49)NG,IX,N(1),M(1),N(2),N(3),M(2),M(3),N(4),N(5),N(6),N(7),
    1M(4),M(5),M(6),M(7)
    IF(NG)1,1,300
  300 NG=NG-1
    IF(NG-3)100,100,7
C LEE TATARABUELOS DE IX
  7 READ(5,50)(N(I),I=8,15),(M(I),I=8,15)

```

PARCON - EFN SOURCE STATEMENT - IFN(S) -

```

      IF(NG-4)100,100,8
C LEE RETATARABUELOS DE IX
      8 READ(5,50)(N(I),I=16,31)
      READ(5,50)(M(I),I=16,31)
C LEE ANTEPASADOS CONSANGUINEOS DE M(1) O DE IX Y SUS COEFICIENTES DE
C CONSANGUINIDAD
      100 READ(5,51)(NAC(K),K=1,14)
      READ(5,52)(CA(K),K=1,14)
      600 IF(NG-1)58,76,75
C LIMITES DEL NUMERO DE ANTEPASADOS A COMPARAR DE N(1) Y M(1)
      75 IF(NG-5)11,12,77
      11 LI=14
      LJ=14
      GO TO 13
      12 LI=30
      LJ=30
      GO TO 13
      77 LI=62
      LJ=62
      GO TO 13
C LIMITES DEL NUMERO DE ANTEPASADOS A COMPARAR DE LA VIA PATERNA Y
C MATERNA DE IX
      76 IF(NG-4)10,11,12
      10 LI=6
      LJ=6
C DETECCION DE ANTEPASADOS COMUNES
      13 LI=LI
      LJ=LJ
      IC=2
      JC=2
      PAR=0.
      KON=0
      14 I=IC
      140 I1=I
      J=JC
      150 J1=J
      IF(N(I)-M(J))36,15,36
      15 IF(N(I)+M(J))16,36,16
C DETECCION DE ARBORESCENCIAS GENEALOGICAS
      16 IF(I-1)26,26,17
      17 I=I/2
      IF(J-1)23,23,18
      18 J=J/2
      IF(N(I)-M(J))20,19,20
      19 IF(N(I)+M(J))24,20,24
      20 IF(J-1)24,21,18
      21 IF(I-1)24,22,25
      22 I=I1
      J=J1
      GO TO 26
      23 I=IC
      GO TO 26
      24 I=I1
      J=J1
      GO TO 36
      25 J=J1

```

```

GO TO 17
C VALORACION DE ARBORESCENCIAS GENEALOGICAS
26 COEF=1.
27 IF(I-1)29,29,28
28 I=I/2
   COEF=COEF/2.
   GO TO 27
29 IF(J-1)31,31,30
30 J=J/2
   COEF=COEF/2.
   GO TO 29
31 I=I1
   J=J1
   DO 33 K=1,28
   IF(NAC(K))32,34,32
32 IF(N(I)-NAC(K))33,35,33
33 CONTINUE
34 PAR=PAR+COEF
   GO TO 36
C FACTOR DE CORRECCION DE CONSANGUINIDAD
35 S=COEF+(COEF*CA(K))
C OBTENCION DEL COEFICIENTE DE PARENTESCO
   PAR=PAR+S
   S=0.
36 J=J+2
   IF(J-LJ)150,150,37
37 I=I+2
   IF(I-LI)140,140,90
C AJUSTE DE LIMITES DEL NUMERO DE ANTEPASADOS DEL MISMO SEXO A COMPARAR
90 KON=KON+1
   IF(KON-2)38,39,40
38 IC=1
   JC=1
   LI=LI+1
   LJ=LJ+1
   GO TO 14
39 IC=1
   JC=2
   LIR=LI-1
   LI=1
   LJ=LJ-1
   GO TO 14
40 IF(KON-3)42,41,42
41 IC=2
   JC=1
   LI=LIR
   LJ=1
   GO TO 14
42 IF(NC-1)58,44,43
C ESCRIBE LA PAREJA N(1),M(1), NUMERO DE GENERACIONES Y COEFICIENTE DE
C PARENTESCO
43 WRITE(6,53)N(1),M(1),NG,PAR
   GO TO 2
44 NG=NG+1
C OBTENCION DEL COEFICIENTE DE CONSANGUINIDAD
   CONS=PAR/2.

```

PARCON - EFN SOURCE STATEMENT - IFN(S) -

```

C ESCRIBE EL INDIVIDUO IX, NUMERO DE GENERACIONES Y COEFICIENTE DE
C CONSANGUINIDAD
   WRITE(6,54)IX,NG,CONS
   GO TO 6
58 STOP
END

```

