

INTÉRÊT DE LA NUMÉRATION BINAIRE DANS LES RECHERCHES GÉNÉALOGIQUES

J. LEFEBVRE

*Station centrale de Génétique animale,
Centre national de Recherches zootechniques, Jouy-en-Josas (Seine-et-Oise)*

SOMMAIRE

L'étude, dans toute leur complexité, des liaisons entre les caractéristiques d'un animal et celles de l'ensemble de ses ascendants, exige d'une part l'emploi des méthodes statistiques à variables multiples. Ceci est devenu possible grâce à la rapidité et à la capacité des calculateurs actuellement disponibles. La nature binaire du langage interne de ces appareils se révèle d'autre part particulièrement bien adaptée à la structure des généalogies. La note présente quelques remarques permettant de simplifier et d'accélérer considérablement l'établissement et l'édition des généalogies, ainsi que leur utilisation pour les calculs de consanguinité.

Depuis près d'un siècle, les animaux domestiques ont fait l'objet d'améliorations remarquables, qu'il s'agisse de leur conformation ou de performances aussi différentes que la vitesse des chevaux de course ou la production des vaches laitières.

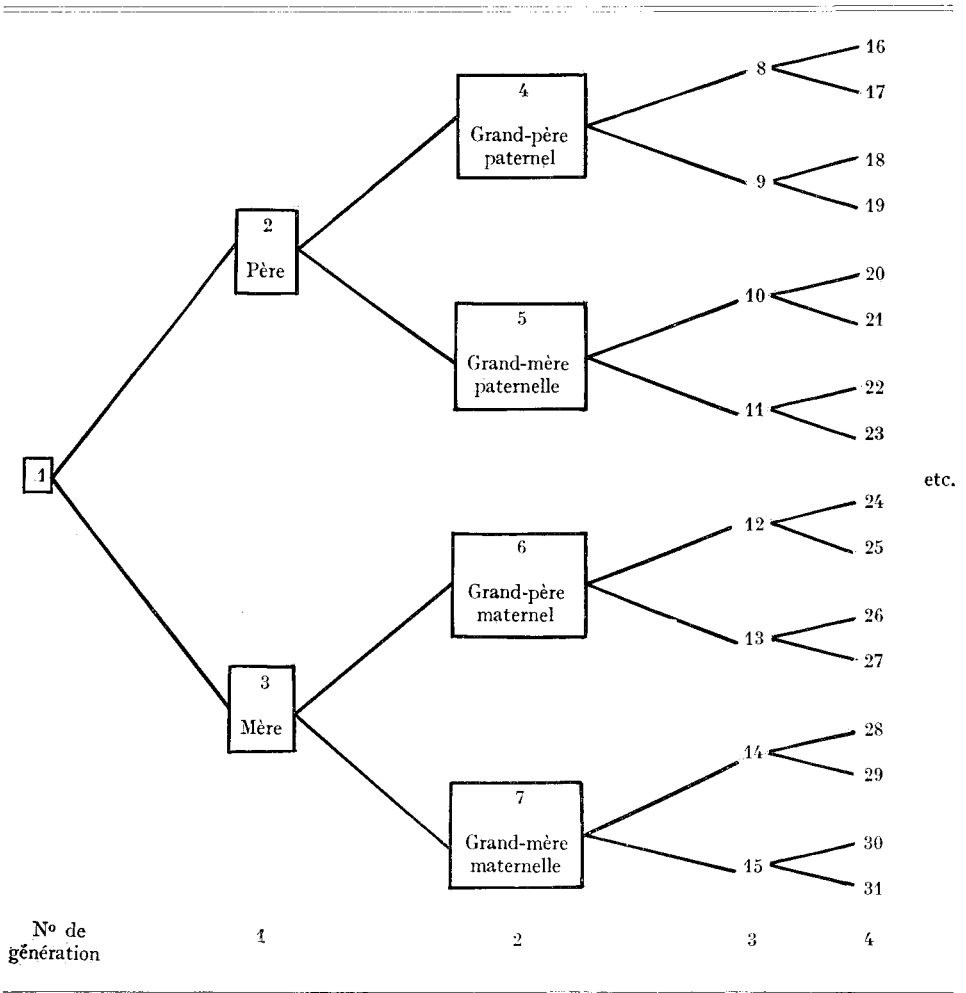
Les origines des races actuellement les plus répandues dans le monde sont généralement bien connues et elles se limitent à un très petit nombre de reproducteurs exceptionnels, remarquables et utilisés efficacement par un nombre également réduit de sélectionneurs. Le ralentissement du rythme des progrès génétiques depuis quelques décennies, les échecs rencontrés parfois lors de la transposition à l'échelle de populations nombreuses, inséminées artificiellement, de méthodes qui s'étaient révélées efficaces au stade d'élevages limités et homogènes méritent qu'on approfondisse les raisons de cette situation. Pour en juger valablement, il est nécessaire de rechercher, dans toute leur complexité, les relations qui peuvent exister entre caractéristiques d'ascendants et de descendants.

Jusqu'à présent, les études ont été limitées aux éléments de deux générations successives. Or, un sélectionneur n'a jamais choisi un reproducteur uniquement en

fonction des productions de sa mère par exemple, ou même des produits déjà connus de son père. Il prend en considération, d'une manière très intuitive, les défauts et qualités de ses ascendants et collatéraux, sur de nombreuses générations.

Pour qu'une telle synthèse devienne moins empirique et ne reste pas le fait de quelques personnes douées de dons d'observation et d'une mémoire exceptionnels, deux problèmes doivent être résolus : pouvoir manier rapidement des généalogies comportant un grand nombre de générations, étudier de façon systématique les relations entre les caractéristiques de production des sujets constituant ces généalogies.

TABLEAU I



Les calculateurs actuels, par leur rapidité et leur capacité, permettent d'utiliser les méthodes statistiques à variables multiples, indispensables pour l'analyse simultanée de ces liaisons. De plus leur mode de calcul binaire est bien adapté à l'établisse-

ment des généalogies. Considérons en effet le système de numérotage des ascendants qui a été proposé à plusieurs reprises, mais dont l'emploi ne s'est pas généralisé, car ses avantages sont très limités dans le cas d'un traitement manuel (tabl. 1).

Sous cette forme décimale, on voit immédiatement que les numéros de rangement des « têtes de générations », 1, 2, 4, 8, 16 ... peuvent s'exprimer en puissances successives de 2 : 2^0 , 2^1 , 2^2 , 2^3 , 2^4 , ..., dont les exposants sont justement les numéros d'ordre des générations. Tous les numéros des éléments d'une même génération contiennent la même puissance de 2. La distance, exprimée en nombre de générations, qui sépare deux éléments de la généalogie, et qui est utilisée dans les calculs de parenté et de consanguinité, est donc donnée directement par la différence entre les puissances de 2 contenues dans les numéros de ces deux éléments.

Si nous traduisons, comme le calculateur dans son langage interne, ces numéros en binaire, nous obtenons le schéma du tableau n° 2.

Nous pouvons, sur ce nouveau schéma, faire les remarques suivantes :

1° Le nombre de chiffres, moins un, du numéro d'ordre d'un antécédent est égal au numéro d'ordre de sa génération. La division par 2 en numération décimale est donc remplacée en binaire par la détermination de la position du dernier chiffre de gauche, opération 4 à 5 fois plus rapide.

2° Le dernier chiffre du numéro d'ordre d'un ascendant donne immédiatement son sexe : 0 pour les mâles, 1 pour les femelles. La recherche de la parité ou de l'imparité, succession de 2,5 comparaisons en moyenne, est remplacée par la comparaison d'une seule position binaire, 2 à 3 fois plus rapide.

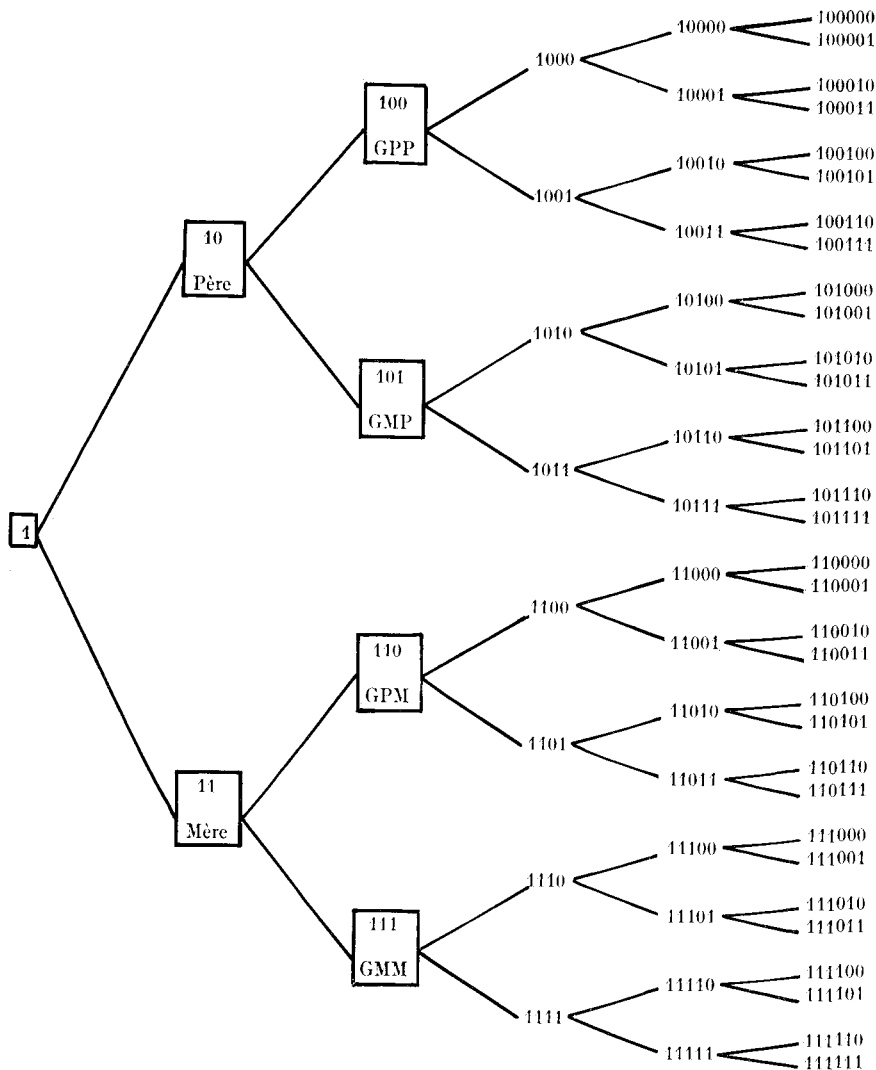
3° Les numéros de tous les ascendants d'un animal commencent tous par le numéro de cet animal.

4° Tous les numéros des éléments de la lignée mâle commencent par 10, tous ceux de la lignée femelle par 11.

La constitution d'une généalogie est une opération simple : si l'on dispose, pour chaque individu d'une population donnée, d'un élément (carte perforée par exemple), où, à son indicatif, (numéro d'inscription, nom à condition qu'il soit unique...), sont associés ceux de ses parents, on « enregistre » sur l'une des mémoires du calculateur ces éléments, triés pour accélérer les opérations ultérieures. Sur une seconde mémoire, on range, en les numérotant respectivement 1, 2, 3, l'indicatif de l'individu dont on désire la généalogie et ceux de ses parents. La recherche de l'indicatif du père dans la première mémoire, donne les indicatifs des grands-parents paternels que l'on range dans la seconde mémoire en les numérotant 4 et 5. On opère de même pour les grands-parents maternels et, successivement, pour les éléments des générations successives. La généalogie est donc constituée, par colonnes successives, en ajoutant à chaque indicatif (immuable et caractéristique de chaque individu), un (ou plusieurs) numéros d'ordre, propres à la généalogie constituée, et qui y fixent son (ou ses) emplacements. Mais la généalogie constituée dans la seconde mémoire ne comporte pas de « blanc » (dans le cas d'antécédents inconnus, ce qui arrive fréquemment lorsque le nombre des générations est élevé) : tous les antécédents sont rangés à la suite les uns des autres, leur place dans la généalogie étant donnée par leur numéro d'ordre et non par leur emplacement dans la mémoire.

Le nombre d'éléments d'une généalogie croît rapidement avec le nombre de générations (tabl. 3).

TABLEAU 2



N° de génération	1	2	3	4	5
------------------	---	---	---	---	---

TABLEAU 3

Nombre de générations.	1	2	3	4	5	6	7	8	9	10
Nombre d'éléments de la dernière colonne.....	2	4	8	16	32	64	128	256	512	1 024
Nombre total d'éléments.	3	7	15	31	63	127	255	511	1 023	2 047

Si la constitution d'une généalogie est simple, elle est donc toujours relativement longue au-delà de 4 ou de 5 générations. Sa longueur dépend également de l'importance de la population de base.

Il y a donc intérêt à pouvoir reprendre directement les ascendances déjà constituées, pour les animaux qui se retrouvent fréquemment. Cette réutilisation revient à un changement des numéros d'ordre. Ce changement, qui n'apparaît guère possible en numération décimale, se résume par contre à une simple substitution des premiers chiffres en numération binaire.

Prenons, par exemple, l'antécédent 100 dans le tableau 2 et supposons que nous le retrouvions, dans la même généalogie, ou dans une autre, en 110. Ses propres ascendants dans la première généalogie (tabl. 4) doivent devenir pour 110 ceux du tableau n° 5.

TABLEAU 4

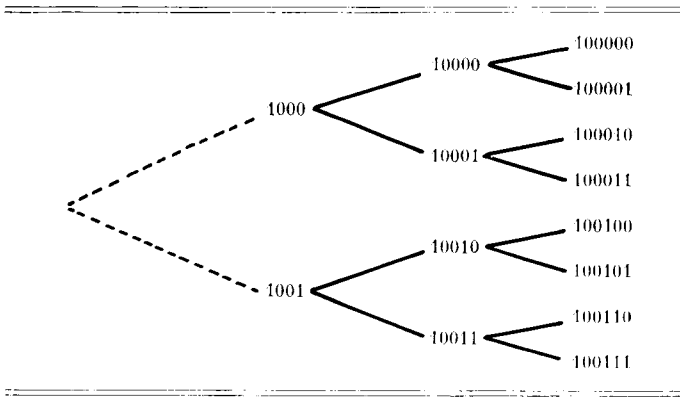
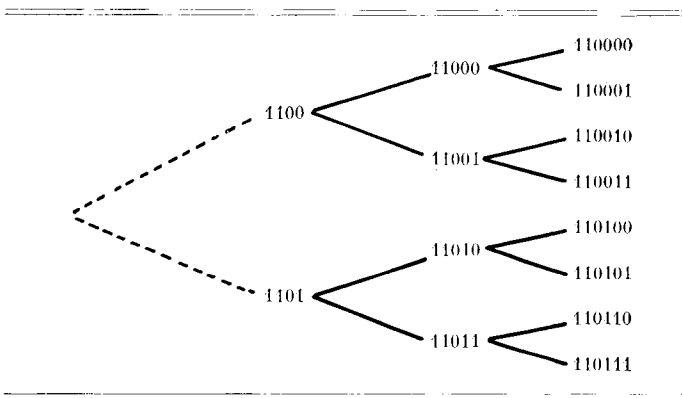


TABLEAU 5



On voit immédiatement que le passage d'un tableau à l'autre revient, dans chaque numéro d'ordre, à remplacer les 3 premiers chiffres 100 par 110.

Si le transfert doit s'effectuer d'une génération à une supérieure, l'opération est

très semblable. Si par exemple, 100 se retrouve en 1 100, il suffira de remplacer les 3 premiers chiffres (100) de chaque numéro d'ordre par 1 100.

Un cas particulier intéressant de transfert à une génération d'ordre inférieur est fourni par la mise en réserve, sur cartes par exemple, de la généalogie d'un animal que l'on pense rencontrer souvent dans les recherches généalogiques conduites sur une population déterminée. Il est naturel de lui affecter le numéro 1 pour ce stockage. Si l'on veut conserver l'origine du III, on substituera simplement 1 aux 3 premiers chiffres de chaque antécédent, qui sont toujours III.

Les gains de temps réalisables grâce aux remarques précédentes, dépendent étroitement des calculateurs et de la nature des mémoires de stockage. Mais, quels qu'ils soient, les transferts indiqués avec un simple changement de quelques chiffres du numéro d'ordre, sont beaucoup plus rapides que le mode de recherche, à partir de la population de base, des ascendants.

TABLEAU 6

En numération décimale						En numération binaire											
1	←	2	←	4	←	16	1	←	10	←	100	←	1000	←	10000		
						17									10001		
						18							1001	←	10010		
						19									10011		
				5	←	10	←	20					101	←	1010	←	10100
						21									10101		
						22									10110		
						23									10111		
						24									11000		
				3	←	6	←	12	←	24					11001		
						25									11010		
						26									11011		
						27									11100		
						28									11101		
				7	←	14	←	28					111	←	1110	←	11110
						29									11111		
						30									11110		
						31									11111		

La numérotation binaire accélère, en second lieu, l'édition des généalogies. La présentation verticale « en arbre », habituelle aux généalogistes, ou celle « horizontale » couramment utilisée en zootechnie, avec des flèches ou accolades reliant chaque animal à ses deux parents, peuvent difficilement être envisagées avec les « sorties » dont les calculateurs disposent généralement. Par contre, une légère modification de la présentation « horizontale » conduit à une solution compatible avec l'organisation du matériel existant. Elle consiste à sortir sur la première ligne la succession des têtes de générations et dans la dernière colonne, les éléments successifs de la dernière génération. Chaque ligne est alors constituée de droite à gauche, à partir des éléments de cette colonne, en divisant par 2 leur numéro d'ordre jusqu'à obtenir un nombre impair, ce qui en binaire, revient à supprimer, dans le numéro d'ordre, le chiffre de droite jusqu'à ce qu'il soit 1 (tabl. 6).

La numération binaire entraîne également de grandes simplifications dans les calculs de parenté et de consanguinité. Ces calculs commencent par la recherche de la présence simultanée d'un même antécédent dans les origines de 2 individus, les parents dans le cas de consanguinité.

La première remarque donne immédiatement la « distance » en générations.

La deuxième remarque, sur le sexe, permet de réduire de moitié le nombre de comparaisons sur les indicatifs, en effectuant au préalable une comparaison sur la dernière position du numéro binaire seulement. Dans le cas d'indicatifs de 10 chiffres, le temps de recherche est réduit de 40 p. 100.

La troisième remarque permet, dès qu'un animal a été trouvé simultanément dans les lignées mâle et femelle, de garder mémoire (sous forme du numéro de ses deux positions) de la « voie » empruntée afin de ne plus l'utiliser ultérieurement. Ce problème ne semble pas avoir été résolu en numération décimale.

La quatrième remarque permet, partant d'un antécédent de la ligne 10, d'éviter toute autre comparaison d'indicatif dans cette lignée, par une comparaison préalable à 10 des 2 premiers chiffres du numéro d'ordre. La réduction totale de temps passe à 55 p. 100.

Il est probable que d'autres propriétés de la numération binaire seraient susceptibles de faciliter encore davantage les recherches généalogiques sur calculateur électronique. Mais dès à présent, il nous a paru utile d'indiquer celles qui nous ont permis d'aborder certaines études que nous comptons développer prochainement.

Reçu pour publication en mai 1965.

SUMMARY

ADVANTAGE OF THE BINARY NUMERATION IN PEDIGREE RESEARCHES

The study of the very complex links existing between the characters of an animal and those of its ascendants first demands the use of multivariate analysis. This is now possible thanks to the new speed and efficiency of modern computers. On the other hand, the binary numeration of these computers is found to be fairly well adapted to the structure of pedigrees. This paper contains a few remarks that enable considerably quicker and simpler establishment and use of pedigrees in consanguinity computation.

I. N. R. A.
BIBLIOTHEQUE UD 3590
DOMAINE DE CROUELLE
63039
CLERMONT-FD CEDEX 2