

Logiciel d'analyse de généalogies adapté à de grandes populations

Didier BOICHARD

Dernière mise à jour août 2015

Le logiciel PEDIG, développé pour l'analyse de grandes populations, est un ensemble de programmes indépendants mais écrits de façon homogène en Fortran et selon une philosophie commune, en vue de caractériser la qualité des généalogies disponibles, de calculer des probabilités d'origine de gènes et des coefficients de parenté et de consanguinité.

Journal des versions

Nature des mises à jour de la V3:

- Ajout des programmes `verif_ped.f`, `contribution.f`, `anc_comm.f`

Nature des mises à jour de la V4-V5

- Passage des sources en allocation dynamique
- Executables pour linux et windows
- Modification de `ped_util` pour la codification des individus

Contenu du zip

Ce fichier téléchargé comprend 6 répertoires :

- `doc` : les documentations françaises et anglaises de PEDIG, ainsi que différents articles rapportant des résultats produits avec lui.
- `Source` : les fichiers fortran source, complets et autonomes
- `Bin` : les exécutables pour Windows (Dos)
- `Bin_linux` : les exécutables pour Lunix

Les exécutables sont produits avec gfortran.

- `ex1`, `ex2`, `ex3`, `ex4` : exemples réunissant les fichiers d'entrée, de paramètres, de log et de sorties des programmes.

Sous Windows et Linux, les exécutables du répertoire `bin` sont directement utilisables. Ils prévoient une allocation dynamique de sorte que les dimensions sont adaptées à la taille de l'application.

Pour d'autres systèmes ou en cas de besoin de recompilation, un compilateur fortran est nécessaire. Pour Windows, le compilateur libre gfortran est disponible à l'adresse : <https://gcc.gnu.org/wiki/GFortranBinaries>

Citation : Pedig doit être cité comme suit :

Boichard D., 2002. Pedig : a fortran package for pedigree analysis suited to large populations. *7th World Congress on Genetics Applied to Livestock Production*, Montpellier, 19-23 août 2002, paper 28-13.

Présentation générale des programmes

Préparation du fichier de pedigree nécessaire aux autres programmes

Le programme *ped_util* extrait le pedigree complet d'une liste d'individus d'intérêt à partir d'un fichier ressource éventuellement plus large, où chaque individu est codé avec son numéro d'origine. Le programme :

- sélectionne les individus d'intérêt et remonte itérativement tous leurs ancêtres, sur toutes ou un nombre maximum de générations
- élimine, optionnellement, les ancêtres sans parents connus et avec un seul descendant sélectionné dans la sous-population
- recode ce pedigree sélectionné en séquence, de sorte qu'il soit utilisable par les autres programmes. Trois ordres sont disponibles : selon l'ordre des identifiants des animaux ('N'), les parents avant les produits ('T') ou les produits avant les parents ('R')
- sauvegarde, optionnellement, d'autres informations comme l'année de naissance ou le sexe, informations utilisées par les autres programmes.

Un exemple est fourni dans le répertoire 'ex3'. Le fichier sortie 'geneal' contient tous les ancêtres des individus présents dans le fichier entrée 'reference', en se basant sur le fichier de pedigree complet 'original_geneal'.

Le programme *ped_util* suppose que les animaux sont identifiés par un numéro à 14 caractères maximum, que toutes les informations sont séparées par un séparateur, et qu'un parent inconnu est codé '0'. Le code du parent inconnu et le séparateur sont paramétrables.

Vérification de la généalogie

Le programme *verif_ped.f* vérifie que la généalogie est possible. Une généalogie impossible correspond au fait qu'un individu se retrouve dans sa propre généalogie. Si le programme indique une anomalie, il réalise diverses recherches d'erreurs et propose des corrections dans le fichier sortie.

Un exemple est fourni dans le répertoire 'ex4'.

Caractérisation de l'information généalogique

Le programme *ngen.f* calcule par sexe et année de naissance le nombre moyen d'ancêtres connus, le pourcentage d'ancêtres connus à chaque génération, ainsi que le nombre d'équivalent-génération remonté, défini comme la somme sur toutes les générations de la proportion d'ancêtres connus par génération. Il constitue un indicateur simple et synthétique de la qualité de l'information généalogique. Ces statistiques ne sont calculées que pour les individus ayant au moins un parent connu. La connaissance des pedigrees individuels peut être sauvegardée en sortie, sur 12 générations. Le nombre de générations remontées est paramétrable.

Le programme *anc_comm.f* recherche des ancêtres communs à un ensemble (de relativement petite taille) d'individus et des parents de ces individus. Il détermine les ancêtres communs à la plupart d'entre eux. Pour chaque ancêtre commun aux parents des individus recherchés, il produit le fichier de généalogie réduit aux relations nécessaires entre cet ancêtre et les aux individus analysés. Le nombre de fichiers en sortie dépend du nombre d'ancêtres communs trouvés. Ce programme est particulièrement utile pour représenter ensuite un arbre généalogique conduisant à un ancêtre commun à un groupe d'animaux (par exemple, présentant tous la même anomalie récessive).

Le programme *intgen.f* calcule les intervalles de génération sur les quatre voies père-fils, mère-fils, père-fille et mère-fille, en utilisant les années de naissance des animaux. En sortie, il fournit optionnellement l'âge des parents des mâles, pour l'analyse éventuelle des individus d'intérêt (mâles d'IA par exemple).

Le programme *intgend.f* calcule les intervalles de génération (en jours) sur les quatre voies père-fils, mère_fils, père-fille, mère_fille, en utilisant les dates de naissance des animaux sous le format *yyyymmdd*. Seules les dates connues et complètes sont utilisées. La date de naissance est lue à la position fournie par l'utilisation (position égale ou supérieure à 6, les 5 premières étant l'individu, le père, la mère, l'année de naissance et le sexe). Seules les dates comprises entre le 1^{er} janvier 1950 (19500101) et le 31 décembre 2020 (20201231) sont utilisées.

Consanguinité

Deux programmes sont disponibles pour calculer les coefficients de consanguinité de tous les individus d'une population.

Le programme *meuw.f*, utilise la méthode de Meuwissen et Luo (1992), une variante de l'approche de Quaas (1976). Cette méthode est fondée sur la construction du facteur de Cholesky de la matrice de parenté. L'originalité de la méthode de Meuwissen et Luo est de construire ce facteur de Cholesky ligne par ligne. Chaque ligne est construite en remontant tout le pedigree de chaque individu. Le coefficient de consanguinité de l'individu est ensuite déduit des éléments de cette ligne et des coefficients de consanguinité des ancêtres. Cette méthode est la plus rapide actuellement disponible, du moins pour les pedigree les plus courants d'une longueur inférieure à 15-20 générations. Donc dans le cas général, ce programme très efficace est à privilégier, l'autre (*vanrad.f*) étant réservé aux situations plus spécifiques (présence de groupes de fondateurs ou nombre de générations fixé).

Un autre programme, *vanrad.f*, utilise la méthode de VanRaden (1992), une variante de la méthode tabulaire. Le principe consiste à construire la matrice de parenté d'un individu et de tous ses ancêtres, le coefficient de consanguinité étant ensuite déduit de la diagonale de cette matrice. Trois fois moins rapide que la méthode de Meuwissen et Luo, cette méthode peut être utilisée pour a) calculer les coefficients de consanguinité à nombre de générations borné et ainsi de dissocier consanguinité proche et consanguinité éloignée ; b) considérer que les ancêtres sont potentiellement apparentés entre eux ou consanguins, hypothèse utile lors de

l'analyse de populations dont l'information généalogique est hétérogène. A noter que ces deux options sont incompatibles entre elles.

Ces deux programmes fournissent en sortie les coefficients de consanguinité des individus consanguins.

Le programme *grain.f* (Baumung et al, 2015) calcule des consanguinités ancestrales et partielles par simulation et « gene dropping ». Différents indicateurs sont disponibles :

(1) le coefficient de consanguinité classique défini par la probabilité que les deux copies d'un gène tiré au hasard soient identiques par descendance.

(2) Le coefficient de consanguinité ancestrale (f_{a_BAL}) selon la définition de Ballou (1997) avec la formule itérative suivante : $f_a = [f_{a(s)} + (1 - f_{a(s)})f_s + f_{a(d)} + (1 - f_{a(d)})f_d]/2$, où f_a et f sont la consanguinité ancestrale et f la consanguinité classique. L'indice s ou d représente ces valeurs pour le père et la mère. Dans cette formule, f_a et f sont supposés indépendants. Quand ce n'est pas le cas, f_a est surestimé.

(3) Le coefficient de consanguinité ancestrale (f_{a_KAL}) selon la définition de Kalinowski et al (2000). Le coefficient de consanguinité standard est séparé en deux parties, selon que les allèles ont subi la consanguinité dans le passé (consanguinité ancestrale, f_{a_KAL}) ou la subissent pour la première fois (consanguinité nouvelle, f_{new}).

(4) Le coefficient de consanguinité ancestrale (f_{a_BAU}) selon la définition de Baumung. Le concept s'appuie sur le fait que les allèles ont subi la consanguinité plus souvent dans le passé ont une plus forte probabilité de ne pas être délétères que ceux qui ne l'ont pas connu.

(5) Les coefficients de consanguinité partiels. C'est la probabilité qu'un individu soit homozygote pour un allèle IBD d'un fondateur spécifique. La somme sur tous les fondateurs pour un individu est le coefficient de consanguinité total (Lacy et al. 1996; Lacy 1997).

Probabilités d'origine de gènes

Le programme *prob_orig.f* calcule les probabilités d'origine de gènes d'un ensemble d'individus (défini par exemple par leur sexe et leur période de naissance). Les probabilités d'origine de gènes sont combinées entre elles pour estimer un nombre de fondateurs efficaces. Le fichier sortie fournit les contributions non nulles de chaque ancêtre (fondateurs ou non), ainsi que les contributions de chaque ancêtre. A noter que la somme des contributions des fondateurs est 1, tandis que la somme des contributions des ancêtres est supérieure, du fait des redondances.

Optionnellement, ce programme recherche itérativement les ancêtres (pas forcément fondateurs) les plus importants et en déduit un nombre d'ancêtres efficaces. Le nombre d'ancêtres efficaces, défini par Boichard *et al* (1995, 1997), est analogue au nombre de fondateurs efficaces mais il est calculé à partir des contributions marginales des ancêtres les plus importants. Il présente l'avantage de prendre en compte les goulots d'étranglement existant dans les pedigree, et d'être moins sensible à une connaissance imparfaite ou hétérogène des pedigree.

Enfin, ce programme fournit en sortie le fichier '*list_ancestors*' qui peut être utile en entrée du programme *grain.f* ou *contribution.f*.

Le programme *contribution.f* détermine la contribution cumulée d'une série d'ancêtres donnés à une population de référence définie par une liste ou par le sexe et les années de naissance.

Le programme *segreg.f* simule la ségrégation des gènes des fondateurs dans la population pour estimer le nombre de génomes de fondateurs efficaces encore présents dans la population actuelle, à partir des fréquences géniques observées.

Une application fréquente des probabilités d'origine de gènes est le calcul de la contribution d'un nombre limité d'origines à une population, ainsi que l'évolution de cette contribution au cours du temps. Une application pratique, par exemple, est le suivi de la contribution d'une race originelle dans un croisement d'absorption ou une lignée synthétique. Ce travail peut être réalisé avec le programme *prob_orig*. Toutefois, l'analyse de l'évolution au cours du temps nécessiterait de lancer ce programme un grand nombre de fois. Le programme *etr.f* réalise cette analyse en un seul passage. Il se base sur une classification des individus fondateurs, qui doit être fournie dans le fichier de pedigree. Optionnellement, lorsque l'origine n'est pas connue avec certitude, du fait d'une connaissance généalogique insuffisante, des valeurs par défaut de la composition génétique dépendant du sexe et de l'année de naissance peuvent être introduites. Si ces données sont omises, un individu sans parent connu est supposé dériver de l'origine 1. En sortie, la composition génétique individuelle est fournie.

Coefficients de parenté

Les programmes *par3.f* et *parente.f* calculent la matrice de parenté complète d'un ou de deux ensembles d'individus prédéterminés. Le mode de calcul est très différent dans ces deux programmes.

Dans *parente.f*, la liste des individus et de tous leurs ancêtres est constituée à partir du fichier de pedigree, puis la matrice de parenté correspondante est construite par la méthode tabulaire. La partie correspondant aux individus sélectionnés est ensuite sauvegardée. Si deux groupes sont définis, l'édition isole la parenté entre groupes de la parenté intra groupe. **Cette méthode n'est adaptée qu'à un nombre d'individus choisis restreint à quelques centaines au maximum.**

Le programme *par3.f* calcule la matrice de parenté complète terme par terme, en créant pour chaque couple un produit fictif puis en calculant son taux de consanguinité. Contrairement à *parente.f*, *par3.f* **n'a pas de limitation particulière en mémoire**. Cependant, il peut être long en temps de calcul (l'application dans 'ex3' demande d'environ 2 mn de calcul pour environ 10 millions de coefficients).

Pour mesurer le niveau de parenté moyen dans une grande population, le programme *par.f* réalise des calculs de parenté sur des échantillons tirés au hasard et répétés, suivant différentes classifications possibles : échantillons aléatoires dans

l'ensemble de la population née durant une période donnée, échantillon intra troupeau ou intra département (dérivé du numéro du troupeau), échantillons aléatoires de couples. Le processus est le suivant : détermination de la population candidate dans le fichier de pedigree, tirage d'un échantillon, calcul de sa matrice de parenté par la méthode tabulaire, calcul des statistiques correspondantes, réplication autant de fois que nécessaire sur d'autres échantillons.

Le programme *par2.f*, proche conceptuellement de *par3.f*, estime la parenté moyenne entre des individus d'une part et un groupe de référence d'autre part. Ce groupe de référence peut être défini à partir du fichier de généalogie avec des critères de sexe et d'année de naissance, ou à partir d'une liste. Ce groupe peut être exhaustif ou échantillonné. Contrairement à *par.f*, ce programme crée un produit pour chaque accouplement considéré et calcule le coefficient de consanguinité du produit avec la méthode de Meuwissen. Une application possible de ce programme est l'analyse de la parenté entre un ensemble de mâles et la population femelle ou un noyau de sélection.

Principe de fonctionnement du logiciel

Tous ces programmes acceptent en entrée le même fichier de pedigree, qui contient les informations suivantes (même si ces informations ne sont pas toutes utilisées par chaque programme) :

- ⇒ numéro d'individu, en séquence de 1 à N. Le fichier est trié suivant ce numéro.
- ⇒ numéro du père. Un père inconnu est codé 0 ou par un nombre négatif, correspondant alors à un groupe, dans la méthode de VanRaden
- ⇒ numéro de la mère. Même remarque que pour le père.
- ⇒ Année de naissance (exemple 1989). Cette donnée sert à la définition des populations de référence (exemple : analyse des probabilités d'origine de gènes des femelles nées de 1992 à 1994). Cette donnée n'est pas utilisée pour trier les individus, de sorte que des incohérences entre parent et produit sont éventuellement acceptées. Si l'année n'est codée que sur deux chiffres, une année inférieure à 20 est recodée en ajoutant 2000 tandis qu'une année supérieure ou égale à 20 est recodée en ajoutant 1900. Pour le programme *intgend*, l'année est remplacée par la date (8 caractères, aaaammjj).
- ⇒ Sexe : Les mâles sont codés 1, les femelles 2. Une autre codification, utile pour distinguer certains animaux particulier, est acceptée le plus souvent mais elle est incompatible avec les programmes *etr.f*, *ngen.f* et *par.f*.
- ⇒ Groupe d'origine, nécessaire uniquement pour le programme *etr.f*.
- ⇒ Elevage, utilisé uniquement par le programme *par.f* pour calculer les parentés intra troupeau ou intra département.

Il est donc conseillé de construire un fichier unique contenant ces 7 informations dans cet ordre, ce fichier pouvant alors être utilisé par tous les programmes. Le programme *ped_util* peut être utilisé pour cette étape de création.

Bien que les algorithmes utilisés nécessitent un pedigree ordonné avec les parents avant leurs produits, **cette contrainte n'est pas requise dans le fichier** de pedigree. En effet, tous les programmes réalisent ce tri (très rapide) en étape préliminaire. Par contre, les identificateurs originaux ne sont pas supportés, les individus doivent être renumérotés en séquence et le fichier de généalogie doit être trié selon ce numéro. Ped_util est donc une option pratique pour préparer les fichiers.

Chaque programme utilise un fichier de paramètres simple, contenant les informations suivantes, et éventuellement un fichier auxiliaire.

programme *ped_util.f*

`nom du fichier de généalogie d'origine'
'nom du fichier des individus analysés'
'nom du fichier sortie'
nombre maximum de générations
nombre d'informations supplémentaires à conserver
élimination des branches mortes (y/n)
optionnel : code du parent inconnu (0 par défaut)
optionnel : code du séparateur du fichier lu (blanc ' ' par défaut)
optionnel : ordre de numérotation (par défaut, 'N' = ordre des numéros, 'T' = les parents avant les produits, 'R' = les produits avant les parents)

Le fichier de généalogie d'origine est supposé contenir les triplets individu-père-mère, codés sur 14 caractères maximum et séparés d'un blanc (ou d'un autre séparateur autre que la tabulation), suivis des informations supplémentaires (sous forme d'entiers, comme l'année de naissance, le sexe, le groupe, etc). Dans le fichier sortie, on trouve un triplet individu-père-mère pour tous les individus demandés et pour chacun de leurs leurs ascendants. Le fichier sortie contient les numéros recodifiés de l'individu, du père et de la mère, les informations supplémentaires, puis les numéros d'origine. En sortie, les parents inconnus sont codés 0.

programme *verif_ped.f*

`nom du fichier de généalogie'

programme *anc_comm.f*

`nom du fichier de généalogie'
'nom du fichier des individus à caractériser'
'nom du fichier sortie'

programme *ngen.f*

`nom du fichier de généalogie'
'nom du fichier sortie (no si pas de fichier)'. Ce fichier contient pour chaque individu le pourcentage d'ancêtres connus sur les 8 premières générations.

Optionnel : nombre de générations caractérisées. Cette option est utile si le nombre de générations devient élevé et limitant compte tenu de la mémoire disponible.

programme *intgen.f*

`nom du fichier de généalogie'

programme *intgend.f*

`nom du fichier de généalogie'

programmes *etr.f*

`nom du fichier de généalogie'

`nom du fichier sortie' ('no' si pas de fichier). Ce fichier contient pour chaque individu sa composition génétique, si la contribution du groupe 1 n'est pas de 100%.

optionnellement, pour chaque année de naissance : l'année, le génotype moyen des mâles (en pour mille, autant de valeurs que d'origines), puis le génotype moyen des femelles. Si une année est omise, le génotype est supposé de l'origine 1.

Exemple pour trois origines :

1975 400 400 200 700 0 300

programme *meuw.f*

`nom du fichier de généalogie'

`nom du fichier sortie contenant les coefficients de consanguinité'

programme *vanrad.f*

`nom du fichier de généalogie'

`nom du fichier sortie contenant les coefficients de consanguinité'

nombre de générations considérées (0=toutes)

pour chaque groupe de fondateurs, la parenté moyenne intra groupe (optionnel, uniquement si 0 à la ligne précédente)

programme *grain.f*

`nom du fichier de généalogie'

Nombre de simulations

Option demandée (0 à 5), voir ci dessous

`nom du fichier des ancêtres à analyser (*list_ancestors*, fichier fourni par *prob_orig*, est une possibilité)'

Les options 0 à 5 sont les suivantes : 1 = consanguinité totale, 2 = consanguinités totale et partielle, 3 = Fab totale et partielle, 4 = Fac totale et partielle, 5 = Fak totale et partielle, 0 = les 5 fichiers sorties.

Si le fichier des ancêtres à analyser est celui produit par *prob_orig.f*, le nom du fichier est '*list_ancestors*'.

programme *prob_orig.f*

`nom du fichier de généalogie'

`nom du fichier sortie contenant les contributions brutes des fondateurs et des autres ancêtres, ou 'no' si ces contributions ne sont pas demandées

nombre d'ancêtres importants recherchés, sexe, années de naissance début et fin de la population de référence considérée

programme *contribution.f*

`nom du fichier de généalogie'

`nom du fichier des ancêtres à caractériser'
`nom du fichier de la population de référence' ou 'no' si définition par le sexe et les années de naissance

Si 'no' à la question précédente, sexe, années de naissance début et fin de la population de référence considérée

programme *segreg.f*

`nom du fichier de généalogie'
nombre de simulations réalisées, sexe, années de naissance début et fin de la population de référence considérée

programme *parente.f* et *par3.f*

`nom du fichier de généalogie'
'nom du fichier des individus à caractériser'
'nom du fichier sortie'

Le programme (*parente* ou *par3*) demande un fichier contenant la liste des individus à caractériser, ainsi que leur groupe d'affectation. Les individus sont répartis dans un groupe codé 1, ou dans deux groupes codés 1 et 2. Le fichier sortie contient une ligne par terme de la matrice de parenté. Si le fichier est nommé 'no', il n'est pas créé et le programme ne fournit que les statistiques de parenté intra et entre groupes.

programme *par.f*

`nom du fichier de généalogie'
option de calcul, années de naissance début et fin de la population de référence, nombre d'échantillons, taille des échantillons. L'option -1 fournit une aide.

programme *par2.f*

`nom du fichier de généalogie'
'nom du fichier des individus à analyser'
'nom du fichier du groupe de référence' (no si pas de fichier)
si le paramètre précédent est 'no', taille (0=tout le groupe), sexe et années de naissance début et fin de la population de référence

Par2 demande donc un fichier d'individus à analyser, et le listing fournit un résultat pour chaque individu de ce fichier. Il demande optionnellement un fichier contenant les individus à comparer.

Compilation

La version V5 est en allocation dynamique : les fichiers sont lus pour en connaître la taille. Il n'est plus nécessaire de définir des tailles maximales et le fichier *blk.incl* des versions précédentes n'est plus utilisé.

Avant la compilation des programmes, choisir la langue des sorties : pour le français, recopier d'abord le fichier *format_fr.incl* sous le nom *format.incl* ; pour l'anglais, recopier d'abord le fichier *format_en.incl* sous le nom *format.incl*.

Sous programmes externes.

Les programmes *segreg*, *par*, *par2* et *grain* réalisent des tirages aléatoires. Les sous-routines correspondantes sont celles du logiciel **RANLIB**, de Brown et Lovato.

Les programmes *par* et *ped_util* réalisent des tris internes et utilisent les sous-routines de **SLATEX** (Jones et al).

Ces routines externes sont mises à disposition librement par leurs auteurs et sont fournies dans le logiciel.

Applications

La plupart de ces programmes ont été développés pour l'analyse généalogique des populations bovines. Différentes applications sont présentées dans Maignel (1995), Boichard *et al* (1995, 1996, 1997) et Moureaux *et al* (2000). Ils sont caractérisés avant tout par leur grande rapidité. Le programme *par3* peut être long si le nombre de coefficients de parenté à calculer est très élevé.

Exemples

Des exemples sont fournis dans les répertoires *ex1*, *ex2*, *ex3* et *ex4*. Un exemple de généalogie impossible est fourni dans *ex4*. *Ex1* correspond à un petit exemple de taille très limitée. *Ex2* correspond à l'analyse d'une population bovine de taille 12 000 environ. *Ex3* fournit aussi un exemple d'utilisation de *ped_util*.

Dans ces exemples, le fichier de pedigree s'appelle geneal. Les paramètres sont fournis dans les fichiers *p_xxx* (où *xxx* est le nom du programme), les sorties correspondantes dans *s_xxx*, et les fichiers de résultats dans *res_xxx*.

exemple d'utilisation : `meuw <p_meuw >s_meuw`

Dans les répertoires exemples, *execut.bat* réalise les analyses correspondants aux exemples.

Enfin, quelques exemples à plus grande échelle, extraits du travail de L. Maignel, sont donnés en annexe.

Références

Baumung R., Farkas J., Boichard D., Mészáros G., Sölkner J., Curik I. (2015). Grain: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *Journal of Animal Breeding and Genetics*, 132, 100–108

Boichard D., Maignel L., et Verrier E. (1995). Estimation of the effective number of founders of a population from pedigree information. 2nd European Workshop on Advanced Biometrical Methods in Animal Breeding, Salzburg, Austria, June 12-20 1995

Boichard D., Maignel L., et Verrier E. (1996). Analyse généalogique des races bovines laitières françaises. *INRA Prod. Anim.* 9, 323-335

Boichard D., Maignel L., et Verrier E. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol* 29, 5-23

Maignel L. (1995)Analyse de la variabilité génétique des bovins laitiers français sur la base de l'information généalogique. Mémoire DAA INAPG, Paris, 70pp.

Meuwissen T.H.E. et Luo Z. (1992) Computing inbreeding coefficient in large populations. *Genet. Sel. Evol.* 24, 305-313

Quaas R.L. (1976) Computing diagonal elements of a large numerator relationship matrix. *Biometrics* 92, 949-953

VanRaden P.M. (1992) Accounting for inbreeding and crossbreeding in genetic evaluation for large populations. *J. Dairy Sci.* 75, 3136-3144.

Ce logiciel est disponible librement à l'adresse

<http://www.jouy.inra.fr/gabi/Outils-scientifiques/Logiciels/Pedig>

En cas de problème important ou de bug, contacter : didier.boichard@jouy.inra.fr

Annexe

Programme PED UTIL

date and time Tue Aug 4 15:36:08 2015

Length of animal Id : 14
Code of unknown parents : 0

Pedigree file :original_geneal
File of reference animals : reference
Output file :geneal
Maximum Nb of generations traced : 100
Number of parameters to be read and written : 4
Elimination of useless pedigree : n
Code of unknown parents : 0000000000
Delimiter of input file : ' '
Animal numbering option : N

Memory required for allocation (Gb) : 1

First records of the pedigree file

1	AT00002026	0000000000	0000000000	1950	1	3	0
2	AT00002180	0000000000	0000000000	1950	1	3	0
3	AT00002503	0000000000	0000000000	1950	1	3	0
4	AT00002920	0000000000	0000000000	1950	1	3	0
5	AT00003130	0000000000	0000000000	1950	1	3	0
6	AT00003170	0000000000	0000000000	1950	1	3	0
7	AT00003210	0000000000	0000000000	1950	1	3	0
8	AT00003815	0000000000	0000000000	1950	1	3	0
9	AT00004420	0000000000	0000000000	1950	1	3	0
10	AT18121789	0000000000	0000000000	1950	1	3	0

Nb of individuals in the pedigree file : 108979
Nb of individuals in the list : 4412

Selection of ancestors

At iteration 1, 4447 ancestors selected and 18 generation Nr updated
At iteration 2, 3459 ancestors selected and 27 generation Nr updated
At iteration 3, 2541 ancestors selected and 22 generation Nr updated
At iteration 4, 1959 ancestors selected and 19 generation Nr updated
At iteration 5, 1394 ancestors selected and 15 generation Nr updated
At iteration 6, 850 ancestors selected and 11 generation Nr updated
At iteration 7, 369 ancestors selected and 7 generation Nr updated
At iteration 8, 130 ancestors selected and 0 generation Nr updated
At iteration 9, 40 ancestors selected and 0 generation Nr updated
At iteration 10, 9 ancestors selected and 0 generation Nr updated
At iteration 11, 0 ancestors selected and 0 generation Nr updated

Output format : (3i10,4i10,3(1x,a14),1x,4i3)

First records of the output file

1	0	0	1950	2	1	0	000000002B	0000000000	0000000000
2	0	0	1950	2	1	0	0000000312	0000000000	0000000000
3	0	0	1950	1	1	0	0000000344	0000000000	0000000000
4	0	0	1950	1	1	0	0000000602	0000000000	0000000000
5	0	0	1969	1	1	0	0000000809	0000000000	0000000000

19610 records written in the file geneal
END of SELECTION

Programme VERIF PED

Date et heure de debut : Sun Nov 7 13:58:32 2010

VERIFICATION DU PEDIGREE

Entrez le nom du fichier de pedigree
Fichier lu :geneal

Nombre d individus : 2351
On cherche a ordonner le pedigree par iterations
en commençant par les fondateurs puis
en descendant le pedigree

It	/ Classés	Total / a l	It /	Restent a classer
1	86	44		2265
2	970	884		1381
3	1814	844		537
4	1979	165		372
5	2020	41		331
6	2020	0		331
7	2020	0		331

Ce pedigree est impossible
Recherche des causes possibles

Phase 1 : Edition des animaux non classés
avec descendants (400 premiers)

Individu	Pere	Mere	Descendants
98	1408	1503	1
247	1770	1930	2
883	1146	1159	3
891	1144	1177	33

.....
Nombre total non classé avec descendants : 98

Phase 2 : verification simple sur 2 générations

a) On cherche des individus issus d eux memes
0 individu(s) issus d eux memes

b) On cherche des parents issus de leur produit
0 parent(s) issus de leurs produits

Phase 3 : Recherches de boucles complexes

Elimination comme parents
des individus non classés (un par un)
et analyse des consequences sur le pedigree

Ind	Elimine	NDesc	Iter	Corrections
1	98	1	11	5
2	247	2	5	331

Cette correction resout le probleme !.....

Programme NGEN

Calcul par sexe et année de naissance

- * du nombre moyen d'ancêtres par individu
- * du pourcentage d'ancêtres connus par rang d'ascendance
- * du nombre d'équivalent-génération complètes

Année de naissance : 91
Sexe : 2
avec parents connus
Effectif : 2182
Equivalent nbre de générations connues : 3.85
Nb moyen d'ancêtres par individu : 66.02

Génération Proportions

1	1.0000
2	.8498
3	.7263
4	.5873
5	.3726
6	.2024
7	.0855
8	.0246
9	.0050
10	.0006
11	.0001
12	.0000

Programme MEUW

Temps cpu : 1.10

Statistiques de consanguinité

Effectif total	:	106520
Effectif consanguin	:	18500
Consanguinité moyenne (des consanguins)	:	2.487
Consanguinité maximale	:	31.250

Effectif par classe		
0	5	15862.
5	10	2001.
10	15	513.
15	20	49.
20	25	3.
25	30	71.
30	35	1.

Programme PROB ORIG

Nombre de fondateurs importants recherches : 50
 Sexe de la population de reference : 2
 Annee de naissance debut et fin : 88 91

individus lus : 106520
 # non parents : 64674
 Effectif : 9971
 # fondateurs : 6109
 Fondateurs efficaces : 68.8

Id	S	AN	Contributions			Père	Mère	ND	Nb eff d'anc		
			brute	marg	cumul				Mini	Maxi	
1	1937	0	1970	.0960	.0960	.0960	934	7763	1982	11.12	106.91
2	421	0	1963	.0898	.0898	.1859	167	5545	3139	11.93	57.47
3	2158	0	1976	.0830	.0830	.2689	<u>396</u>	26580	2645	13.35	41.20
4	4516	0	1980	.0709	.0709	.3397	1962	84480	1459	13.61	34.16
5	398	0	1963	.0687	.0687	.4085	139	5281	1132	17.02	29.42
6	4196	0	1977	.0842	.0421	.4506	688	69034	2575	17.95	27.97
7	6666	2	1955	.0365	.0365	.4871	225	5516	3	20.33	26.97
8	4727	0	1982	.0239	.0239	.5110	5198	72478	592	21.19	26.56
9	<u>396</u>	0	1962	<u>.0612</u>	<u>.0197</u>	.5307	239	5247	1568	21.80	26.29

Avec 50 ancêtres,

Minorant : 25.18
 Majorant : 25.34

Programme PAR

Statistiques de parente

date et heure de debut Thu Mar 28 16:01:02 1996

Entrez le fichier de genealogies

Fichier lu :/prod2/dbo/r12

Entrer les paramètres suivants

1:Option 2:Année naiss début 3:Année naiss fin 4:Nb échant
5:Taille échant

Option -1 : aide

Option 0 : males

Option 1 : population femelle totale

Option 2 : femelles intra-troupeau

Option 3 : échantillonnage intra-département

Option 4 : échantillonnage par département

OPTION ECHANTILLONNAGE DE LA POPULATION FEMELLE

Population : femelles nees entre annee debut et annee fin

50 echantillons aleatoires de taille 50

Annee de naissance debut : 88
Annee de naissance fin : 91
Nb animaux lus : 106520
Anomalies de genealogies : 0
Nbre de candidates : 9971

Nbre d echantillons : 50.
Taille moyenne des echantillons : 50.0
Nbre de coefficients : 61250.
Moyenne des coefficients : .029
Ecart type des coefficients : .033

Frequence des coefficients

0- 1	:	21184.	35	:	*****
1- 2	:	11862.	19	:	*****
2- 3	:	7736.	13	:	*****
3- 4	:	6504.	11	:	*****
4- 5	:	3618.	6	:	*****
5- 6	:	1824.	3	:	***
6- 7	:	2293.	4	:	****
7- 8	:	1646.	3	:	***
8- 9	:	867.	1	:	*
9- 10	:	373.	1	:	*
10- 11	:	151.	0	:	
11- 12	:	62.	0	:	
12- 13	:	1424.	2	:	**
13- 14	:	912.	1	:	*
14- 15	:	501.	1	:	*
15- 16	:	168.	0	:	
16- 17	:	91.	0	:	
17- 18	:	16.	0	:	
18- 19	:	11.	0	:	
19- 20	:	5.	0	:	
20- 21	:	1.	0	:	
25- 26	:	1.	0	:	

date et heure de fin Thu Mar 28 16:02:33 1996

OPTION ECHANTILLONNAGE FEMELLE INTRA TROUPEAU

Tous les troupeaux sont consideres

Annee de naissance debut : 88

Annee de naissance fin : 91

Nb animaux lus : 106520

Anomalies de genealogies : 0

Nbre d elevages : 656

Nbre de candidates : 9579

Nbre d echantillons : 656.

Taille moyenne des echantillons : 14.6

Nbre de coefficients : 117388.

Moyenne des coefficients : .042

Ecart type des coefficients : .045

Frequence des coefficients

0- 1	:	28412.	24	:	*****
1- 2	:	18739.	16	:	*****
2- 3	:	14604.	12	:	*****
3- 4	:	13627.	12	:	*****
4- 5	:	8943.	8	:	*****
5- 6	:	5422.	5	:	*****
6- 7	:	5602.	5	:	*****
7- 8	:	4406.	4	:	****
8- 9	:	2703.	2	:	**
9- 10	:	1475.	1	:	*
10- 11	:	828.	1	:	*
11- 12	:	354.	0	:	
12- 13	:	3965.	3	:	***
13- 14	:	2823.	2	:	**
14- 15	:	2010.	2	:	**
15- 16	:	1229.	1	:	*
16- 17	:	723.	1	:	*
17- 18	:	312.	0	:	
18- 19	:	217.	0	:	
19- 20	:	125.	0	:	
20- 21	:	62.	0	:	
21- 22	:	19.	0	:	
22- 23	:	14.	0	:	
23- 24	:	3.	0	:	
24- 25	:	1.	0	:	
25- 26	:	465.	0	:	
26- 27	:	160.	0	:	
27- 28	:	58.	0	:	
28- 29	:	58.	0	:	
29- 30	:	12.	0	:	
30- 31	:	4.	0	:	
31- 32	:	10.	0	:	
32- 33	:	1.	0	:	
37- 38	:	2.	0	:	