

Apprecier la connexion dans une évaluation génétique internationale : exemple des races Montbeliarde et Simmental

S. MINERY (1), M.-N. FOUILLOUX (1), S. MATTALIA (1), D. LALOË (2)

(1) Institut de l'Élevage, INRA-SGQA - 78352 Jouy-en-Josas

(2) INRA, Station de Génétique Quantitative et Appliquée - 78352 Jouy-en-Josas

RESUME - La connexion génétique entre pays est un élément essentiel de la fiabilité d'une évaluation génétique internationale. Un faible niveau de connexion peut biaiser la comparaison d'individus de pays différents et fausser le classement général des reproducteurs. Cet article présente une nouvelle méthode pour mesurer la connexion : des différences de niveau génétique entre pays sont simulées, puis un BLUP est réalisé utilisant les *pedigree* réels afin de réestimer les différences introduites initialement. Plus deux pays échangent de gènes et plus leurs niveaux génétiques seront proches. D'autre part, plus ces deux pays sont connectés, mieux on réestime les différences initiales. Cette méthode a été testée sur les populations qui participent aux évaluations conjointes Interbull en race Simmental et Montbeliarde : l'Italie est un des pays le mieux connecté, tandis que la Slovénie apparaît comme le pays le moins bien connecté. L'Irlande et les Pays-Bas utilisent principalement des taureaux Montbéliards français. Cette méthode est très simple et rapide à mettre en application en routine.

Assessment of connectedness in the international genetic evaluation of Simmental and Montbeliard breeds

S. MINERY (1), M.-N. FOUILLOUX (1), S. MATTALIA (1), D. LALOË (2)

(1) Institut de l'Élevage, INRA-SGQA - 78352 Jouy-en-Josas

SUMMARY - Connectedness between countries is a critical point for a reliable international evaluation. Weak genetic links between countries can affect predicted genetic values and ranking of the animals. A new method is presented here to measure connectedness between countries: a systematic difference between true genetic levels of countries is introduced by simulation, a BLUP is performed based on true pedigree, then the percentage of the initial difference that can be found between the estimated genetic levels of the corresponding countries is measured. The more 2 countries exchange genes, the closer are their true genetic levels. The better 2 countries are connected, the higher is the percentage of true re-estimated genetic difference. It was then shown that among countries involved in the Interbull Simmental and Montbeliard breed evaluation, Italy is one of the best connected countries and Slovenia probably the worst connected one. Ireland and the Netherlands use only Montbeliard French bulls. This method was very easy and rapid to implement and could be used in routine.

INTRODUCTION

Tout l'intérêt du BLUP réside en la comparaison de valeurs génétiques entre animaux ou groupes d'animaux (troupeaux, pays...). La fiabilité de ces comparaisons est étroitement liée aux connexions génétiques existant entre ces groupes. Dans l'évaluation internationale des bovins laitiers (Interbull), la connexion entre pays est actuellement appréciée par le nombre de taureaux communs entre pays (taureaux avec des filles dans plusieurs pays donc des index dans différents pays). Mais ces valeurs ne représentent que très partiellement les différents liens génétiques qui existent entre deux pays. Notamment, le nombre de filles qu'ont ces taureaux communs dans chaque pays n'est pas considéré. Rekaya *et al.* (1999) ont introduit la "similarité génétique" représentant la proportion de filles issues des taureaux communs dans chaque paire de pays. Des méthodes quantitatives ont également été proposées pour évaluer le degré de connexion en utilisant un coefficient de détermination (CD) (Laloë, 1993 et Laloë *et al.*, 1996, Laloë et Phocas, 2003, Fouilloux et Laloë, 2006). Mais ces méthodes ne sont pas parfaitement adaptées à la situation internationale en raison de problèmes théoriques ou de limites informatiques. Cet article présente une nouvelle méthode apte à quantifier la connexion entre pays dans le contexte des évaluations Interbull. Ce travail intervient dans le cadre de l'action innovante SIMOINT relative au développement des évaluations internationales en race Simmental Française et Montbeliarde.

1. MATERIEL ET METHODES

1.1. THEORIE

La fiabilité d'une évaluation génétique de n animaux peut être analysée par l'examen de $n-1$ comparaisons particulières ou "contrastes canoniques" $\{c_1, \dots, c_i, \dots, c_{n-1}\}$ (Laloë, 1993, Laloë *et al.*, 1996). De fait : 1/ toute comparaison entre animaux est une combinaison linéaire de ces contrastes canoniques : $\mathbf{b} = \sum_{i=1, n-1} a_i c_i$; 2/ le CD de la comparaison est

une moyenne pondérée des CD des contrastes canoniques : $\{c_1, \dots, c_i, \dots, c_{n-1}\}$ $CD(\mathbf{b}) = \frac{\sum_{i=1, n-1} a_i^2 CD(c_i)}{\sum_{i=1, n-1} a_i^2}$;

3/ l'espérance de la valeur prédite de la comparaison s'écrit : $E(\hat{\mathbf{b}}) = \sum_{i=1, n-1} a_i CD(c_i) E(c_i)$. Donc, si $E(c_i)$ n'est pas nul, le

biais de la prédiction de la comparaison est fonction de $CD(c_i)$: $E(\mathbf{b} - \hat{\mathbf{b}}) = \sum_{i=1, n-1} a_i (1 - CD(c_i)) E(c_i)$

Ainsi, la comparaison des valeurs génétiques de 2 taureaux (x et y) de 2 pays (1 et 2) est une somme de comparaisons intra et entre pays. Dans le cas simple d'un plan équilibré où n taureaux sont contrôlés sur descendance dans N pays (s taureaux "nationaux" avec n_p produits et t taureaux "internationaux" avec n_q produits dans chaque pays), seuls 2 types de contrastes et de CD canoniques interviennent dans les comparaisons entre animaux (Laloë et Phocas, 2003) (tableau 1) : les comparaisons de taureaux d'un même pays (v_w et CD_w), et les comparaisons des niveaux génétiques des différents pays (v_b et CD_b).

Ainsi :

$$u_{1x} - u_{2y} = [u_{1x} - \bar{u}_1] - [u_{2y} - \bar{u}_2] + [\bar{u}_1 - \bar{u}_2]$$

$$= \sqrt{\frac{s-1}{s}} v_{w1} - \sqrt{\frac{s-1}{s}} v_{w2} + \sqrt{\frac{2}{s}} v_{b1.2}$$

D'où :

$$CD(u_{1x} - u_{2y}) = \frac{s-1}{s} CD_w + \frac{1}{s} CD_b \quad [1]$$

$$E[(u_{1x} - u_{2y}) - (\hat{u}_{1x} - \hat{u}_{2y})] =$$

$$(1 - CD_w)[E(u_{1x} - \bar{u}_1) - E(u_{2y} - \bar{u}_2)] \quad [2]$$

$$+ (1 - CD_b)E(\bar{u}_1 - \bar{u}_2)$$

Donc, une "disconnexion" entre deux pays, qui se traduit par un CD_b proche de 0, a deux conséquences sur la comparaison de deux taureaux issus de ces mêmes pays :

- La précision (CD) décroît [formule 1], mais de façon négligeable dès que le nombre de taureaux (s) dépasse quelques dizaines,
- Un biais est probable si les niveaux génétiques des deux pays sont différents [formule 2].

S'appuyant sur cette dernière formule, nous proposons une méthode quantifiant la connexion entre pays par la capacité de prédiction du biais en évaluation génétique.

Tableau 1 : comparaisons canoniques

Type de contraste	Nbre de contrastes	Contraste canonique	CD
Intra-pays	$N(s-1)$	$v_w = \sqrt{\frac{s}{s-1}} [u_{ix} - \bar{u}_i]$	$CD_w = \frac{n_p}{n_p + \lambda}$
Entre pays	$N-1$	$v_b = \sqrt{\frac{s}{2}} [\bar{u}_1 - \bar{u}_2]$	$CD_b = \frac{n_p \eta}{n_p \eta + \lambda}$

$\eta = tn_q / (tn_q + sn_p)$, i.e. la proportion d'animaux issus de taureaux internationaux dans chaque pays.

1.2. PRINCIPE ET PROCEDURE

La méthode proposée consiste à simuler une différence entre les niveaux génétiques vrais des pays, réaliser une indexation BLUP et mesurer le pourcentage de la différence initiale de niveau génétique entre pays que le BLUP permet d'estimer.

1.2.1. Simulation de la différence entre pays

Soient n taureaux nés dans N pays différents. Si on cherche à quantifier le niveau de connexion génétique entre le pays k (référant) et les $N-1$ autres (prétendants), la valeur génétique vraie attribuée à tout taureau né dans k sera a_k , celle attribuée aux autres taureaux sera a_0 . Ainsi, le niveau génétique vrai de chaque pays (j) est

$$\bar{u}_j = 0,5 \left(n_{ij} \times a_k + \sum_{l=1, N}^{l \neq k} n_{lj} \times a_0 \right) / \sum_{l=1, N} n_{lj}$$

filles de taureaux nés dans le pays l dont la lactation est enregistrée dans le pays j . Le niveau génétique des pays ainsi calculé renseigne sur le flux de gènes entre le pays référant (k) et les prétendants. Ainsi, \bar{u}_j tend vers :

- $0,5 \times a_k$ si le pourcentage de vaches contrôlées dans j issues de taureaux nés dans le référant augmente.
- $0,5 \times a_0$ si le pourcentage de vaches contrôlées dans j issues de taureaux nés dans un prétendant augmente.

1.2.2. Evaluation génétique (BLUP)

Les valeurs génétiques des taureaux ont ensuite été estimées par un BLUP mono-caractère sur modèle père en considérant une seule génération de taureaux. Le seul effet fixe considéré était le pays de lactation de la vache et l'héritabilité était de 0,30. Le niveau génétique estimé du pays j est donné par : $\hat{u}_j = 0,5 \times \sum_{i=1}^n n_{ij} \times \hat{u}_i / \sum_{i=1, n} n_{ij}$ où \hat{u}_i est la

valeur génétique estimée du taureau i et n_{ij} , le nombre de vaches issues de i dont la lactation est enregistrée dans j .

1.2.3. Mesure du taux d'estimation de la différence initiale

Disposant des niveaux génétiques vrais et estimés des pays, il est possible de calculer les $N-1$ contrastes vrais et estimés entre le pays référant (k) et chacun des prétendants (j) :

$C_{jk} = (\bar{u}_j - \bar{u}_k)$ et $\hat{C}_{jk} = (\hat{u}_j - \hat{u}_k)$; ainsi que leur ratio $R_{jk} = \hat{C}_{jk} / C_{jk}$. Les $N-1$ ratios ainsi obtenus permettent de classer les $N-1$ prétendants selon leur niveau de connexion avec le référant. Comme R_{jk} n'est pas équivalent à R_{kj} , cette procédure doit être appliquée N fois en changeant le pays référant.

1.3. DONNEES

Les données laitières d'août 2005 de l'évaluation internationale de la race Simmental ont été analysées. La France y participe avec la Simmental Française (FRA) et la Montbéliarde (FRM). Le pays d'origine des taureaux est déterminé à partir de leurs numéros d'identification. Au total 18 599 pères issus de 15 populations ou "pays" ont eu 6 558 980 filles avec des lactations enregistrées dans 10 pays. Les effectifs de taureaux communs sont donnés dans le tableau 2.

Tableau 2 : taureaux communs, évaluation laitière Interbull d'août 2005, races Simmental et Montbéliarde

	CHE	CZE	DEA	FRM	FRA	HUN	IRL	ITA	NLD	SVN
CHE	1662	49	66	93	2	0	17	48	27	9
CZE		1597	157	99	9	8	11	74	19	16
DEA			12285	60	62	19	11	276	18	62
FRM				2475	0	1	30	106	56	15
FRA					207	5	0	36	0	13
HUN						108	0	9	1	7
IRL							34	17	16	2
ITA								796	23	36
NLD									65	1
SVN										362

CHE : Suisse, CZE : République Tchèque, DEA : Allemagne-Autriche, FRM : France Montbéliarde, FRA : France Simmental, HUN : Hongrie, IRL : Irlande, ITA : Italie, NLD : Pays-Bas, SVN : Slovénie.

2. RESULTATS ET DISCUSSION

Lors du test, a_k (référant) était égal à 200 et a_0 (prétendant) à 400. La durée des calculs a été inférieure à 10 minutes de CPU et dépendait essentiellement du degré de convergence exigé lors du BLUP. Le tableau 3 présente les résultats pour les 10 pays qui participent aux évaluations Interbull en race Simmental. Chaque pays est considéré référant tour à tour. Les résultats se lisent horizontalement (Ainsi, sur la première ligne la Suisse est le référant). Trois variables différentes ont été calculées : 1/ Le niveau génétique simulé du pays j (référant ou prétendant, u_j noté **NG**), 2/ La différence entre les vrais niveaux génétiques (C_{jk} noté **DIFF**) et 3/ Le taux de réestimation des différences initiales (R_{jk} en pourcentage).

2.1. LECTURE DES RESULTATS

Niveau génétique du pays (NG) : Sur la diagonale, un NG proche de 100 signifie que le pays utilise peu de taureaux d'origine étrangère. Au contraire, un NG proche de 200 indique que le pays utilise en majorité des taureaux étrangers. Hors de la diagonale, NG donne des indications sur les flux de semence : proche de 200, le prétendant n'utilise quasiment pas de taureaux du référent, tandis que proche de 100, il utilise presque exclusivement des taureaux du référent. Ainsi horizontalement NG indique comment les taureaux du référent sont utilisés par les autres pays (exportation) tandis que verticalement, il indique de quel(s) pays le prétendant importe ses taureaux.

La différence de niveaux génétiques vrais (DIFF) : Eu égard aux valeurs a_k et a_0 du test, DIFF peut varier de 100 (s'il n'y a pas d'échange de gènes entre deux pays) à -100 (si les taureaux du prétendant proviennent tous du référent et inversement). La valeur peut aussi être égale à 0 si deux pays ont des taureaux de même origine (référent/prétendants) dans les mêmes proportions et quelque soit le prétendant.

Cependant ces deux critères (NG et DIFF) ne mesurent pas réellement la connexion, ils rendent compte de la situation en terme d'échange de taureaux entre les pays. Si le pays A utilise beaucoup de taureaux du pays B, mais que par ailleurs ces taureaux ne sont pas utilisés dans le pays B, la connexion entre A et B restera faible. Seul le taux de réestimation des différences initiales mesure le degré de connexion. Plus deux pays sont connectés et plus ce pourcentage est élevé. Ces pourcentages doivent être lus horizontalement et indiquent si la différence de niveau génétique entre le référent et chaque prétendant est correctement estimée.

2.2. ANALYSES DE CAS PARTICULIERS

a) Si on exclut IRL et NLD qui seront traités ultérieurement, ITA apparaît comme le pays qui échange le plus de gènes avec les autres (NG sur la diagonale le plus élevé avec 144,2 et DIFF, ITA comme référent, les plus bas (≈ 55)). Parmi les origines des taureaux étrangers utilisés par ITA se trouvent majoritairement DEA (NG, ITA prétendant, DEA référent = 167,9), et FRM (191,3). Dans une moindre mesure ITA importe aussi des taureaux CHE (199,7), FRA (199,9), CZE (199,9) et SVN (199,9). D'autre part, des taureaux ITA sont utilisés en FRA (NG, ITA référent, FRA prétendant = 198,3), SVN (199,3), CZE (199,9) et CHE (199,9). Ces échanges de taureaux contribuent certainement aux fortes connexions entre ITA et les autres pays, avec des pourcentages de réestimation du biais autour de 80,3 % (un des scores globaux les plus élevés).

b) FRM semble être le pays qui utilise le moins de taureaux étrangers (NG sur la diagonale de 100,2 et DIFF entre 91 et 99). Néanmoins les pourcentages de réestimation du biais ne sont pas si bas, comparés aux autres pays et varient de 59,1 (avec SVN) à 70,2 (avec CHE), (si on exclut IRL et NLD). FRM utilise des taureaux CHE (NG, FRM prétendant, CHE référent = 199,9), CZE (199,9), DEA (199,9), ITA (199,9) et SVN (199,9). FRM n'utilise aucun taureau d'origine FRA (200). D'autre part, tous les autres pays, sauf FRA, utilisent des taureaux FRM (NG horizontaux avec FRM comme référent tous inférieurs à 200 sauf avec FRA).

c) La lecture horizontale des NG renseigne sur le fait qu'aucun pays n'utilise de taureaux HUN ou IRL (à part le pays lui-même). La situation est quasiment la même pour

SVN. Cependant HUN et SVN ne présentent pas le même profil de connexion à l'analyse leurs pourcentages de réestimation du biais respectifs. HUN qui a pourtant moins de taureaux communs avec les autres pays que SVN (tableau 2) est mieux connecté aux autres pays que SVN (ses pourcentages moyens de réestimation sont de l'ordre de 66 % contre 39 % pour SVN - tableau 3). En effet, la comparaison en détail de HUN et SVN (tableau 4) montre que, bien qu'ils aient le même pourcentage de taureaux utilisés dans d'autres pays (18,5 % et 17,4 % respectivement), les taureaux communs HUN ont en proportion plus de filles dans le pays que ceux SVN (34,6 % contre 3,9 %). Et avec un pourcentage moyen de réestimation du biais de 39 %, SVN a le niveau le plus faible de connexion.

Tableau 4 : taureaux communs et nombre de filles pour HUN et SVN

	HUN	SVN
Nb taureaux dans le pays (nb filles)	108 (13499)	362 (83455)
Nb taureaux avec filles que dans le pays (nb filles)	88 = 81,5 % des taureaux (8822 = 65,3 % des filles)	299 = 82,6% des taureaux (80218 = 96,1% des filles)
Nb taureaux avec filles à l'étranger (nb filles)	20 = 18,5% des taureaux (4677 = 34,6% des filles)	63 = 17,4% des taureaux (3237 = 3,9% des filles)

d) NLD et IRL sont dans une situation particulière. Ils ont un NG diagonale très proches de 200 et des DIFF avec chaque prétendant proche de 0, ce qui signifie qu'ils utilisent en très grande majorité des taureaux étrangers. En conséquence en tant que référent, ils ont le même niveau génétique que s'ils étaient prétendants. De plus, lorsque FRM est référent la différence de niveau génétique (DIFF) entre FRM et NLD ou FRM et IRL est égale à 0. En d'autres termes, NLD et IRL utilisent essentiellement des taureaux d'origine FRM. Cela est confirmé par des NG horizontaux de 106,0 et 104,1 pour NLD et IRL. Sur les 34 taureaux IRL, 31 sont originaires de FRM et sur les 65 taureaux NLD, 60 sont issus de FRM. Dans ces cas particuliers, les pourcentages de réestimation du biais sont très difficiles à interpréter du fait des très faibles valeurs des différences de niveau génétique (vrais et estimés).

3. DISCUSSION

L'inconvénient majeur de cette méthode est qu'il n'existe pas un unique critère par paire de pays. Ainsi, le pourcentage de réestimation du biais entre CHE et CZE est 51,9 % lorsque CHE est référent et 58,7 % lorsque CZE est référent. Il faut à chaque fois se placer dans un pays pour analyser la connexion.

Cependant, plusieurs avantages sont clairement identifiés. 1/ Cette procédure est facile à mettre en place, aucune information n'est requise en plus des fichiers de routine fournis par Interbull. 2/ Les trois critères calculés contiennent plus d'information qu'un simple nombre de taureaux communs, sans pour autant multiplier exagérément le nombre de variables. 3/ Enfin, cette méthode mesure la "disconnexion" à travers sa conséquence la plus néfaste : l'impossibilité de comparer le niveau génétique d'animaux de différentes origines ou de différents groupes.

CONCLUSION

Cette méthode est originale et efficace pour analyser les échanges de taureaux entre pays et pour mesurer les connexions génétiques éventuellement créées par ces taureaux communs. Elle peut aider des programmes de testage entre pays, comme EUROTTEST (Dodenhoff *et al.*, 2003) pour la Simmental, à décider quels pays doivent échanger en priorité des doses de semences pour améliorer leurs liens génétiques.

La facilité d'application de cette méthode permettrait de la mettre en place rapidement en routine, afin de compléter l'approche "nombre de taureaux communs" aujourd'hui utilisée pour décrire la connexion entre les pays participant aux évaluations Interbull.

La méthode pourrait être améliorée dans un futur proche. Tout d'abord, il serait intéressant d'intégrer les liaisons génétiques entre pères en ajoutant au moins une génération d'ancêtres pour prendre en compte les liens génétiques indirects entre pays. Par ailleurs, d'autres études pour être

réalisées telles que l'évolution de la connexion dans le temps en Simmental, ou l'analyse de la connexion dans d'autres races impliquées dans les évaluations internationales, comme la Holstein ou la Brune.

Les auteurs remercient Interbull pour les données et le Ministère de l'Agriculture, à travers le projet SIMOINT, pour son soutien financier.

Dodenhoff J., Egger-Danner C., Mattalia S. 2003. *Interbull Bulletin* 31, 137-142

Fouilloux M.N., Laloë D., 2006. *Proc. 8th World Congr. Gen. Appl. Livest. Prod., Belo Horizonte, Brazil*

Laloë D., 1993. *Genet Sel Evol*, 25, 557-576.

Laloë D., Phocas F., Menissier F, 1996. *Genet Sel Evol* 28, 359-378

Laloë D., Phocas F., 2003. *Livestock Production Science* 80(3) 241-256

Rekaya R., Weigel K.A., Gianola D., 1999. *50th Meeting of the EAAP, Zurich, Switzerland, 22-26 August 1999*

Tableau 3 : différence de niveau génétique vrai/estimé entre pays

Référénts	Prétendants										
	CHE	CZE	DEA	FRM	FRA	HUN	IRL	ITA	NLD	SVN	
NG	123,8	199,5	199,8	199,9	198,4	200	200	199,7	196,7	199,8	
CHE	DIFF	75,7	76,0	76,1	74,5	76,2	76,2	75,8	72,9	76,0	
	%	51,9	50,5	58,8	50,2	49,7	58,9	51,6	58,0	49,2	
NG	199,9	116,0	199,9	199,9	200	200	200	199,9	200	199,9	
CZE	DIFF	84,0	84,0	84,0	84,0	84,0	84,0	83,9	84,0	83,9	
	%	58,7	58,6	62,1	58,2	58,6	62,3	59,0	63,5	56,9	
NG	198,2	194,2	115,0	199,9	175,2	169,5	200	167,9	200	188,2	
DEA	DIFF	83,2	79,3	85,0	60,2	54,6	85,0	53,0	85,0	73,3	
	%	56,1	61,4	58,7	84,7	73,6	59,2	75,8	59,6	54,4	
NG	196,8	194,9	199,7	100,2	200	199,5	104,1	191,3	106,0	199,4	
FRM	DIFF	96,6	94,7	99,5	99,8	99,3	3,9	91,2	5,9	99,2	
	%	70,2	66,4	60,5	60,2	59,7	57,4	63,1	41,1	59,1	
NG	200	199,9	199,9	200	130,3	198,0	200	199,9	200	200	
FRA	DIFF	69,7	69,7	69,7	69,7	67,7	69,7	69,6	69,7	69,7	
	%	80,5	80,6	81,1	80,5	81,1	80,5	81,1	80,5	80,4	
NG	200	200	200	200	200	136,7	200	200	200	200	
HUN	DIFF	63,3	63,3	63,3	63,3	63,3	63,3	63,3	63,3	63,3	
	%	65,9	66,0	66,3	65,9	66,2	65,9	66,2	66,0	65,9	
NG	200	200	200	200	200	200	195,9	200	200	200	
IRL	DIFF	4,1	4,1	4,1	4,1	4,1	4,1	4,1	4,1	4,1	
	%	82,8	82,7	82,6	83,5	82,6	82,6	82,7	83,4	82,5	
NG	199,9	199,9	199,9	199,9	198,3	200	200	144,2	200	199,3	
ITA	DIFF	55,7	55,7	55,8	55,8	54,1	55,8	55,8	55,8	55,1	
	%	79,5	79,6	81,2	80,9	81,4	80,3	81,0	81,1	80,1	
NG	199,4	199,9	199,9	200	200	200	200	200	198,4	200	
NLD	DIFF	1,0	1,5	1,6	1,6	1,6	1,5	1,6	1,6	1,6	
	%	118,2	92,0	89,5	93,4	89,3	88,9	93,5	90,0	88,2	
NG	200	199,9	199,9	199,9	200	200	200	199,9	200	116,8	
SVN	DIFF	83,2	83,1	83,1	83,1	83,2	83,2	83,2	83,1	83,2	
	%	38,6	38,7	38,9	38,7	38,8	38,7	38,8	39,1	38,7	

NG : Niveau génétique vrai issu des simulations

DIFF : Différence de niveaux génétiques vrais entre 2 pays

% : Pourcentage de différence vraie réestimée