Belg. J. Zool. — Volume 125 (1995) — issue 1 — pages 101-107 — Brussels 1995

A COMPARISON OF GENE FLOW ESTIMATES BASED ON PRIVATE ALLELE FREQUENCIES

by

HANS DE WOLF (1), THIERRY BACKELJAU (2),
KURT JORDAENS (1), ROBERTO MEDEIROS (3), PETER VERDYCK (2)
and RON VERHAGEN (1)
(1) Department of Biology, Evolutionary Biology, University of Antwerp (RUCA)
Groenenborgerlaan 171, B-2020 Antwerpen (Belgium)
(2) Royal Belgian Institute of Natural Sciences,
Vautierstraat 29, B-1040 Brussel (Belgium)
(3) Departamento de Biologia, Universidade dos Açores
Rua da Mae de Deus 58, P-9502 Ponta Delgada, Açores (Portugal)

SUMMARY

The frequency of private alleles is often used to assess the amount of gene flow (Nm) between populations, with the equations proposed by SLATKIN (1985b) and by SLATKIN and BARTON (1989). Although these equations express the same relationship, they may yield different estimates of gene flow for the same data. These differences increase with decreasing frequencies of private alleles. Comparisons of Nm estimates, based on different equations can therefore be misleading. It is advisable to use these equations method only to distinguish between $Nm > 1$ and $Nm < 1$.

Key words : Population genetics, gene flow, private alleles.

INTRODUCTION

The amount of gene flow (Nm = Number of migrants) between natural populations is usually estimated by indirect methods based on allele frequency data. One of these methods uses the mean frequency $\overline{P}(1)$ of so called private alleles, *i.e.* alleles that are found in one population only (SLATKIN, 1985a). The likelihood that such alleles are exchanged between populations is indeed related to the degree of migration, for the probability of exchange of private alleles between populations will be very low unless migration is frequent. Gene flow estimates based on $\overline{P}(1)$ rely on a simulation model suggesting that in the case of the stepping stone and island models the logarithm of Nm is approximately linearly related to the logarithm of $\overline{P}(1)$. This relationship was first formulated by a nepperian logarithm

(Slatkin, 1985b) and subsequently by a $\log_{10}$ ébased equation (Slatkin and Barton, 1989). Despite expressing the same relationship, different estimates of gene flow can be obtained when both equations are applied to the same data.

In this paper, both equations are compared and some literature on the calculation of gene flow using private alleles is reviewed.

MATERIAL AND METHODS

Using Slatkin's (1985b) formula, gene flow (Nm) is calculated as follows

$$Nm = Nm_{ref} \cdot \frac{25}{N_{sam}}$$

Where Nm_{ref} is a reference gene flow estimate for an arbitrary theoretical sample size (*i.e.* average number of individuals per population) of $N_{ref}=25$ and where N_{sam} is the actual number of individuals sampled per population. Nm_{ref} (for $N_{ref}=25$) can be calculated using following equation

$$Nm_{ref} = e^{(\ln(\overline{P}(1)) - b)\left(\frac{1}{a}\right)}$$

or

$$\ln(\overline{P}(1)) = a\ln(Nm_{ref}) + b$$

with $a=-0.505$ and $b=-2.440$ and where $\overline{P}(1)$ is the average frequency of private alleles over all populations and loci sampled.

Equivalent to these formulae is Slatkin and Barton's (1989) equation where

$$Nm_{ref} = 10^{(\log_{10}(\overline{P}(1)) - b)\left(\frac{1}{a}\right)}$$

or

$$\log_{10}(\overline{P}(1)) = a\log_{10}(Nm_{ref}) + b$$

With values of $a=-0.49$ and $b=-0.95$ for $N_{ref}=10$; $a=-0.58$ and $b=-1.1$ for $N_{ref}=25$ and $a=-0.61$ and $b=-1.2$ for $N_{ref}=50$

Correction for sample sizes different from 10, 25 or 50 is made as follows :

$$Nm = Nm_{ref} \cdot \frac{N_{ref}}{N_{sam}}$$

The formulae of Slatkin (1985b) and Slatkin and Barton (1989) are graphically compared (Fig. 1). $\overline{P}(1)$ values, ln and $\log_{10}$ transformed, ranging from 0.01 to 0.10 are plotted against corresponding ln and $\log_{10}$ transformed Nm estimates obtained using both equations under the assumption that $N_{ref}=25$. Using $\overline{P}(1)$ and average sample sizes adopted from literature, Nm values are recalculated and compared (Table 1). If F_{ST} values (fixation index measuring the degree of genetic differentiation between subpopulations) are known, Nm values are also estimated according to WRIGHT's (1951) method, which is based on the following equation

$$Nm = \frac{1}{4}(\frac{1}{F_{ST}} - 1)$$

Selected F_{ST} and derived Nm estimates are also given in Table 1.

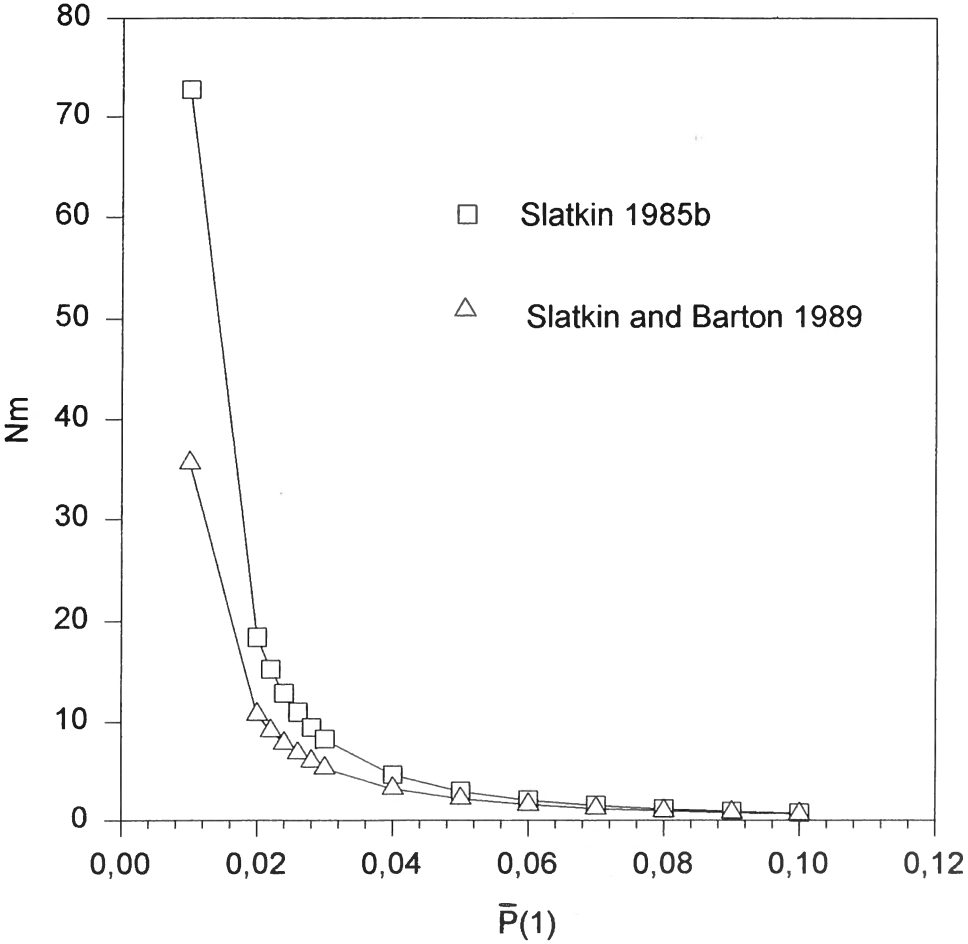

Fig.1. — Graphical comparison of gene flow calculation according to SLATKIN (1985b) and SLATKIN and BARTON (1989) for $N_{sam} = 25$, which are respectively.

$$Nm_{ref} = e^{(\ln(\bar{P}(1)) - b)\,(\frac{1}{a})}$$

$$Nm_{ref} = 10^{(\log_{10}(\bar{P}(1)) - b)\,(\frac{1}{a})}$$

with

$$Nm = Nm_{ref} \cdot \frac{25}{N_{sam}}$$

RESULTS

Fig. 1 shows Nm estimates as function of theoretical $\overline{P}(1)$ values with an average sample size of 25, based on Slatkin's (1985b) and Slatkin and Barton's

TABLE 1

Nm estimates according to Slatkin *(1985b),* Slatkin *and* Barton *(1989) and* Wright *(1951).* N_{sam} *= sample ;* $\overline{P}(1)$ *= mean private alleles frequency ; Nm = gene flow estimate*

Species	N_{sam}	$\overline{P}(1)$	*Nm Slatkin 1985b*	*Nm Slatkin, Barton 1989*	F_{st} *Wright 1951*	*Nm Wright 1951*
Pteridium aquilinum	47.0	.0103	36.51(1)	19.99	0.110	2.022
Pinus ponderosa Deme 157.1	39	.026	7.03	5.39(2)		
Pinus ponderosa Deme 160.5	13	.039	9.45	6.82(2)		
Stephanomeria exigua	45.8	.054	1.4(3)	1.3	0.152	1.394
Mytilus edulis	67.4	.008	42.0(3)	21.1	0.006	41.41
Strombus gigas		.026	10.97[4]	6.85(4)		8.7
Haliotis rubra	90.27	.0071	39.17	19.67(5)		
Haliotis laevigata	72.37	.0106	22.27	14.01(5)		
Gammarus fossarum	33.31	.190	0.16(6)	0.16	0.68	0.117
Gammarus pulex	21.06	.110	0.76(6)	0.67	0.36	0.444
Speonomus hydrophhilus	70	.042	12.53(7)	1.39	0.112	1.982
			1.51			
Sitobion avenae 9 sites	±165	.027	1.6(8)	1.22		
Sitobion avenae 13 sites	±165	.019	3.0(8)	2.16		
Drosophila willistoni	94	.014	9.9(3)	6.0		
Drosophila pseudoobscura	33.2	.075	1.0(3)	0.8	0.200	8.70
Chanos chanos	48.9	.030	4.2(3)	3.3	0.056	4.21
Salmo salar	56	.030	7.7(9)	3.02		
			3.68			
Batrachoseps campi	10.6	.338	0.16(3)	0.09		
Batrachoseps pacifica spp. 1	21.7	.117	0.64(3)	0.59	0.281	0.64
Batrachoseps pacifica spp. 2	22.2	.207	0.20(3)	0.21	0.556	0.20
Hyla regilla	20.1	.081	1.4(3)	1.2		
Plethodon ouachitae	31.4	.054	2.1(3)	1.5	0.106	2.108
Plethodon cinereus	22.1	.200	0.22(3)	0.23		
Plethodon dorsalis	22.3	.294	0.10(3)	0.11		
Lacerta melisellensis	22.4	.066	1.9(3)	1.5		
Peromyscus polionotus	25.2	.158	0.31(3)	0.30	0.446	0.31
Peromyscus californicus	20.0	.066	2.2(3)	1.6		
Thomomys bottae	29.1	.087	0.86(3)	0.73		

[1]Wolf *et al.*, 1991 ; [2]Alstad *et al.*, 1991 ; [3]Slatkin, 1985b ; [4]Mitton *et al.*, 1989 ; [5]Brown and Murray, 1992 ; [6]Scheepmaker, 1990 ; [7]Crouau-Roy, 1989 ; [8]Loxdale, 1990 ; [9]Elo, 1993.

(1989) equations. When $\overline{P}(1)$ values are high, both curves overlap, however as soon as $\overline{P}(1)$ decreases, both curves begin to diverge with the ln based curve (SLATKIN, 1985b) rapidly exceeding the $\log_{10}$ based one (SLATKIN and BARTON, 1989). When $\overline{P}(1)$ reaches for instance 0.01, gene flow estimates drop from 70 to 35 respectively when SLATKIN'S (1985b) or SLATKIN and BARTON'S (1989) equation is used.

The difference between both equations observed in our simulation model is also observed in the literature (Table 1). When Nm values are high (small $\overline{P}(1)$ values), both equations yield substantially different Nm estimates. Nm estimates calculated with SLATKIN (1985b) will then exceed Nm values obtained with SLATKIN and BARTON'S (1989) equation.

This is for example the case in the blue mussel *Mytilus edulis* where the original Nm value drops from 42.0 (see SLATKIN, 1985b) to 21.1 when SLATKIN and BARTON'S (1989) equation is used instead. Similar differences are observed in the bracken *Pteridium aquilinum*. Gene flow between seven British populations was estimated using SLATKIN'S (1985b) equation (WOLF *et al.*, 1991) and yielded Nm = 36.51. This value drops to 19.9 when estimated with SLATKIN and BARTON'S (1989) equation. In the blacklip abalones *Haliotis rubra* and *H. laevigata*, the original Nm values, which are respectively 19.67 and 14.01 (BROWN and MURRAY, 1992) increase to 39.17 and 22.27 when recalculated using SLATKIN'S (1985b) equation.

Although under realistic conditions the method of WRIGHT (1951) is likely to be more accurate than the private alleles method (SLATKIN and BARTON, 1989), one would expect both methods to yield comparable Nm estimates. This is true except for *Pteridium aquilinum* (WOLF *et al.*, 1991). Since the formulae of SLATKIN (1985b) and/or SLATKIN and BARTON (1989) do not show a consistent pattern of difference compared to WRIGHT'S (1951), there is no obvious way to determine which of both formulas matches best with WRIGHT'S (1951) method.

Besides the differences due to the equation used, differences in Nm estimates can also be the result of sample size correction. If the correction is not made according to SLATKIN (1985b) and SLATKIN and BARTON (1989), Nm values could be over- or underestimated. This is the case when sample size correction is ignored as was done in studies of the queen conch *Strombus gigas* (MITTON *et al.*, 1989) and the fruit fly *Ceratitis capitata* (GASPERI *et al.*, 1991). Depending on the values of N_{sam} and N_{ref}, gene flow estimates will then under- ($N_{ref} > N_{sam}$) or overestimate ($N_{ref} < N_{sam}$) the actual Nm value. The same is true when Nm_{ref} is multiplied by an inverted correction term, which was done in a study of the troglobitic beetle *Speonomus hydrophilus* (CROUAU-ROY, 1989). The original Nm value (obtained by multiplying Nm_{ref} with N_{sam} and dividing it by N_{ref}) drops from 12.53 to 1.51 when appropriately corrected.

DISCUSSION

As demonstrated above, the formulae of SLATKIN (1985b) and SLATKIN and BARTON (1989) can give different results when applied to the same data. Since this is particularly the case when $\overline{P}(1)$ values are small it is obvious that this discrepancy will mostly affect gene flow estimates derived from species with a high dispersal potential and hence a high degree of gene flow and low expected frequency of private alleles.

According to WRIGHT (1931) one immigrant per generation ($Nm = 1$) is sufficient to prevent population differentiation due to random genetic drift. The transition from large to small amounts of population differentiation will not occur abruptly with an Nm value of one, yet $Nm = 1$ is very often used as a decisive limit. If gene flow is expressed in terms of its ability ($Nm > 1$) or disability ($Nm < 1$) to prevent population differentiation due to random drift, both equations will yield comparable results. If on the other hand Nm estimates are used to compare gene flow estimates, both equations can not be used interchangeably and the equation used together with $\overline{P}(1)$ and N_{sam} should be specified. Even when specified, confusion may still persist. In ALSTAD *et al.* (1991) SLATKIN'S (1985b) equation, together with values of a and b belonging to SLATKIN and BARTON'S (1989) equation were described. However, Nm was not estimated with SLATKIN'S (1985b) but with SLATKIN and BARTON'S (1989) equation (Table 1). Furthermore it seems that, even when equations are specified, Nm estimates are sometimes compared regardless of the equation used. This was the case in the blacklip abalone *H. rubra* (BROWN, 1991). The $\log_{10}$ based Nm estimate of 19.67 is compared to the ln based Nm estimate of *Mytilus edulis* (see SLATKIN, 1985b), whereas in fact it should be compared with the recalculated Nm value of 21.1, as shown in Table 1.

Given the fact that there seems to be confusion regarding the use of private alleles to estimate gene flow, it is advisable to use SLATKIN'S method only to distinguish between $Nm > 1$ or $Nm < 1$, with $Nm = 1$ as limit. If Nm estimates are to be compared quantitatively, WRIGHT'S (1951) method, based on the mean F_{ST} value, is more appropriate.

ACKNOWLEDGEMENTS

HDW and KJ are funded by an IWT (Belgium) scholarship, PV is funded by an NFWO (Belgium) scholarship. This work was supported by F.J.B.R.-grants 2.0004.91 and 2.0128.94.

REFERENCES

ALSTAD, N.D., C.S. HOTCHKISS, and W.K. CORBIN (1991) — Gene flow estimates implicate selection as a cause of scale insect population structure. *Evol. Ecol.*, **5** : 88-92.

BROWN, L.D. (1991) — Genetic variation and population structure in the blacklip abalone *Haliotis rubra. Aust. J. Mar. Freshw. Res.*, **42** : 77-90.

BROWN, L.D. and N.D. MURRAY (1992) — Population genetics, gene flow and stock structure in *Haliotis rubra* and *Haliotis laevigata*. In SHEPERD, S.A., TEGNER, M.J. and GUZMANN DEL PROO, S.A. (eds.) *Abalone of the world, Biology, Fisheries and culture.* Proceedings of the 1st International Symposium on Abalone : 24-33.

CROUAU-ROY, B. (1989) — Population studies on an endemic troglobitic beetle : geographical patterns of genetic variation, gene flow and genetic structure compared with morphometric data. *Genetics*, **121** : 571-582.

ELO, K. (1993) — Gene flow and conservation of genetic variation in anadromous Atlantic salmon (*Salmo salar*). *Hereditas*, **119** : 149-159.

GASPERI, G., C.R. GUGLIELMINO, A.R. MALACRIDA and R. MILANI (1991) — Genetic variability and gene flow in geographical populations of *Ceratitis capitata* (Wied.) (medfly). *Heredity*, **67** : 347-356.

LOXDALE, D.H. (1990) — Estimating levels of gene flow between natural populations of cereal aphids (Homoptera : Aphididae). *Bull. Entomol. Res.*, **80** : 331-338.

MITTON, B.J., J.C. BERG Jr, and S.K. ORR (1989) — Population structure, larval dispersal, and gene flow in the Queen Conch, *Strombus gigas*, of the Caribbean. *Biol. Bull.*, **177** : 356-362.

SCHEEPMAKER, M. (1990) — Genetic differentiation and estimated levels of gene flow in members of the *Gammarus pulex*-group (Crustacea, Amphipoda) in western Europe. *Bijdr. Dierk.*, **60** : 3-30.

SLATKIN, M. (1985a) — Gene flow in natural populations. *Annu. Rev. Ecol. Syst.*, **16** : 393-430.

SLATKIN, M. (1985b) — Rare alleles as indicators of gene flow. *Evolution*, **39** : 53-65.

SLATKIN, M. and N.H. BARTON (1989) — A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43** : 1349-1368.

WOLF, G.P., F.L.S.E. SHEFFIELD and H.C. HAUFLER (1991) — Estimates of gene flow, genetic substructure and population heterogeneity in bracken (*Pteridium aquilinum*). *Biol. J. Linn. Soc.*, **42** : 407-423.

WRIGHT, S. (1931) — Evolution in Mendelian populations. *Genetics*, **16** : 97-159.

WRIGHT, S. (1951) — The genetical structure of populations. *Ann. Eugen.*, **15** : 323-354.