

Estimation de la taille efficace : état de l'art...

Frédéric HOSPITAL

INRA-GABI
Génétique Animale et Biologie Intégrative
Jouy en Josas, France
frederic.hospital@jouy.inra.fr

May 12, 2016

Notion de nombre efficace

Nombre **réel** : nombre d'éléments (individus) **hétérogènes**

Nombre **efficace** : nombre d'éléments (individus) **homogènes**
ayant le même comportement selon un **critère**

Exemple :

Nombre efficace d'allèles : nombre d'allèles de même fréquence produisant la même variance que le nombre réel d'allèles.

Notion de nombre efficace

Dans les populations naturelles, tous les individus ne participent pas forcément au processus reproductif, si bien qu'en général l'effectif de la population N qui détermine le rythme de la dérive génétique n'est pas égal à l'effectif de recensement de la population...

Notion de nombre efficace

... On définit donc l'effectif efficace de la population (ou taille efficace) comme l'effectif d'une **population idéale** (de type Wright-Fisher) pour laquelle on aurait une **fluctuation du polymorphisme équivalente** à celle de la population naturelle.

Notion de nombre efficace

... C'est donc le nombre d'individus d'une **population idéale** pour lequel on aurait un **degré de dérive génétique équivalent** à celui de la population réelle. On note ce nombre comme **N_e** .

Tout dépend du critère choisi...

Il y a en fait plusieurs types d'effectif efficace, selon à quel effet de la dérive génétique on s'intéresse.

Changement de la consanguinité de la population (inbreeding effective population size).

Dans une population idéale, la probabilité que 2 gènes soient identiques par ascendance à la génération précédente est égale à $1/(2N)$. Le changement du coefficient de consanguinité d'une génération à l'autre est donc donné par

$$1 - F_{t+1} = \left(1 - \frac{1}{2N}\right) (1 - F_t)$$

Donc, dans une population réelle, l'effectif efficace sera celui qui provoquera un changement de consanguinité de même amplitude que dans une population idéale.

$$N_e^{(F)} = \frac{1}{2 \left(1 - \frac{1 - F_{t+1}}{1 - F_t}\right)}$$

Variance du changement des fréquences alléliques d'une génération à l'autre (variance effective population size).

Dans une population idéale, la variance de la fréquence allélique (p_{t+1}) conditionnelle à celle de la génération précédente (p_t) est donnée par la variance binomiale

$$v = \text{var}(p_{t+1}|p_t) = \frac{p_t(1 - p_t)}{2N}.$$

Si \hat{v} est cette même quantité mesurée dans la population réelle, alors l'effectif efficace de variance sera donné par

$$N_e^{(v)} = \frac{p_t(1 - p_t)}{2\hat{v}}$$

Changement de l'hétérozygotie de la population d'une génération à l'autre (eigenvalue effective population size)

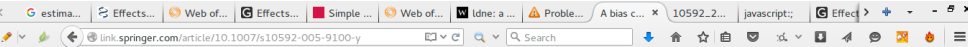
Dans une population idéale, le changement d'hétérozygotie d'une génération à l'autre est donné par

$$H_{t+1} = \left(1 - \frac{1}{2N}\right) H_t$$

Si bien que l'effectif efficace d'une population réelle provoquant le même changement d'hétérozygotie que dans une population idéale sera défini comme

$$N_e^{(H)} = \frac{1}{2 \left(1 - \frac{H_{t+1}}{H_t}\right)}$$

Estimation à partir du déséquilibre de liaison



Download PDF (438 KB)

Article
Conservation Genetics
April 2006, 7:167

First online: 01 March 2006

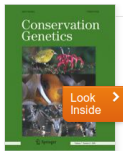
A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*

Robin S. Waples

Download PDF (438 KB)

Abstract

Analysis of linkage disequilibrium (r^2 = mean squared correlation of allele frequencies at different gene loci) provides a means of estimating effective population size (N_e) from a single sample, but this method has seen much less use than the temporal method (which requires at least two samples). It is shown that for realistic numbers of loci and alleles, the linkage disequilibrium method can provide precision comparable to that of the temporal method. However, computer simulations show that estimates of N_e based on r^2 for unlinked, diallelic gene loci are sharply biased downwards ($\hat{N}_e/N_e < 0.1$ in some cases) if sample size (S) is less than true N_e . The bias is shown to arise from inaccuracies in published formula for $E(r^2)$ when S and/or N_e are small. Empirically derived modifications to $E(r^2)$ for two mating systems (random mating and lifetime monogamy) effectively eliminate the bias (residual bias in $\hat{N}_e < 5\%$ in most cases). The modified method also



Reference tools

Export citation

Add to Papers

Other actions

- Register for Journal Updates
- About This Journal
- Reprints and Permissions

Share



academy.springer.com

log in to your
springer.com
account and
learn for free

online
training

Springer

Estimation à partir du déséquilibre de liaison

Applications ▾ Places ▾ iceweasel ▾

ldne: a program for estimating effective population size from data on linkage disequilibrium - WAPLES - 2008 - Molecular Ecology Resources - Wiley Online Library - iceweasel

estima... Effects... Web of... Effects... Simple ... Web of... W ldne... x javascript: W ldne: a ... Effects... Multipl... Service... G E | +

onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2007.02

Search

Most Visited ▾ Google ▾ Google Memotoo Any.do YaZiba Facebook Twitter LinkedIn R W wikipedia Save to Mendeley Linux ▾ Main ▾ Labo ▾

Show messages

	$S \geq 30$	$S < 30$
Random mating		
$E(\hat{p}_{sample}^2)$	$1/S + 3.19/S^2$	$0.0018 + 0.907/S + 4.44/S^2$
\hat{N}_e	$\frac{1/3 + \sqrt{1/9 - 2.76\hat{p}^2}}{2\hat{p}^2}$	$\frac{0.308 + \sqrt{0.308^2 - 2.08\hat{p}^2}}{2\hat{p}^2}$
Monogamy		
$E(\hat{p}_{sample}^2)$	$1/S + 3.19/S^2$	$0.0018 + 0.907/S + 4.44/S^2$
\hat{N}_e	$\frac{2/3 + \sqrt{4/9 - 7.2\hat{p}^2}}{2\hat{p}^2}$	$\frac{0.618 + \sqrt{0.618^2 - 5.24\hat{p}^2}}{2\hat{p}^2}$

N_e is estimated from the overall mean \hat{p}_{Δ}^2 averaged across multiple loci and alleles, which is computed as follows. For each pair of loci i and j , with k_i and k_j alleles, respectively, \hat{p}_{Δ}^2 is computed (equation 1) for each of the $k_i \times k_j$ allelic combinations, and a mean of these allele-pair estimates ($\hat{p}_{\Delta}^2(i, j)$) is calculated for that pair of loci. If L loci are used, there are $L(L-1)/2$ different $\hat{p}_{\Delta}^2(i, j)$ values. Next, two factors are considered in determining the proper weights to give to each $\hat{p}_{\Delta}^2(i, j)$ value in calculating the overall mean \hat{p}_{Δ}^2 : the number of independent alleles and the sample size. Since a locus with k alleles has the equivalent of $k-1$ independent alleles, each $\hat{p}_{\Delta}^2(i, j)$ is based on the equivalent of $n_{ij} = (k_i - 1) \times (k_j - 1)$ independent comparisons. With missing data, the sample size S_{ij} can differ among locus pairs (see next section), and $\hat{p}_{\Delta}^2(i, j)$ values based on larger sample sizes should receive greater weight. LDNE uses weights that are inversely proportional to variances. Hill (1981) provided an approximate formula for the coefficient of variation of \hat{N}_e based on linkage disequilibrium data. For unlinked loci, and using the current notation, this can be rearranged to provide an approximate variance for \hat{N}_e .

URGENT.TXT ... Inbox - Unife... ldne: a progra... R2GA_2016 R2GA_2016... [Estimation d... GMDP Chang... [ProgrammeH... [Multiple esti... 1 / 4

En théorie, (mais pas toujours) les différents types de taille efficace donnent des valeurs identiques. On utilisera une définition ou une autre selon les cas...

Go to old article view

PDF
Info

MOLECULAR ECOLOGY

[Explore this journal >](#)

Original Article

Multiple estimates of effective population size for monitoring a long-lived vertebrate: an application to Yellowstone grizzly bears

Pauline L. Kamath, Mark A. Haroldson, Gordon Luikart, David Paetkau, Craig Whitman, Frank T. van Manen

First published: 28 October 2015 [Full publication history](#)

DOI: 10.1111/mec.13398 [View/save citation](#)

Cited by: 1 article [Refresh](#) [Citing literature](#)

Am scores 52

[Funding information](#)

Abstract

onlineibrary.wiley.com/journal/10.1111/(ISSN)1365-294X



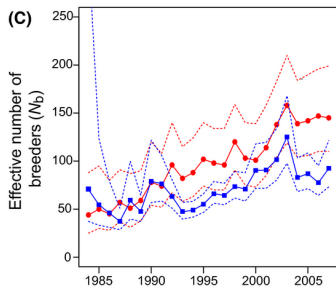
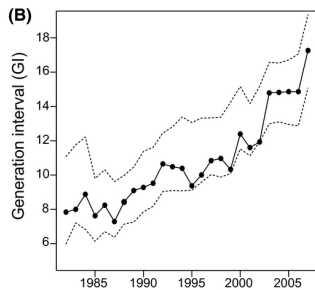
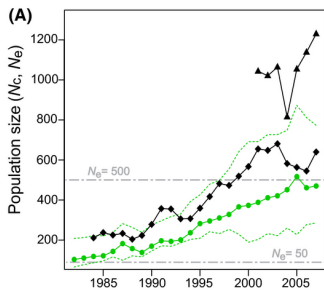
View issue TOC
Volume 24, Issue 22
November 2015
Pages 5507-5521

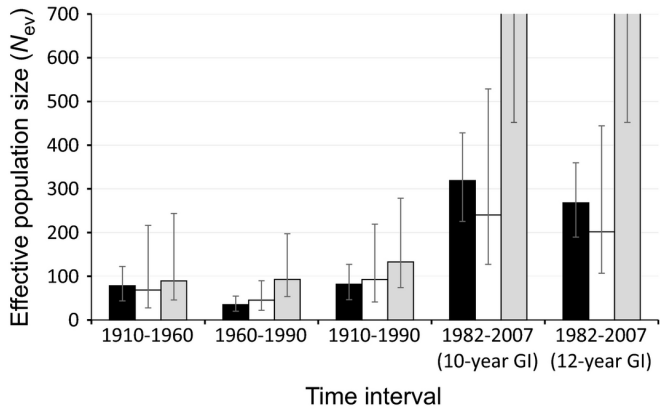
Text size Share

Advertisement

Article Share

Congratulations!
As a Wiley author,
you can share
your article for
FREE with up to
10 of your contacts





Mesure "Intégrées" (coalescence).

Mutation $\Theta = 4 N_e \mu$

Selection $N_e s$

Recombinaison $N_e r$

Mesures de N_e

Multi-échantillons (2 générations ou plus)

- ▶ **Pédigrée**
- ▶ Marqueurs / séquences (H_e , Variance, ...)

Mono-échantillon (1 génération)

- ▶ Marqueurs / séquences (LD, Haplotypes, AFS, Coalescence)

Remarque sur l'estimation à partir du déséquilibre de liaison

- ▶ En théorie, pour des locus liés, mais difficile à implémenter (calculs)
- ▶ A l'origine, application sur des locus indépendants (non-liés). Paradoxe ? Mesure de la variation inter-locus (temps \Rightarrow espace)
- ▶ Développements récents (NGS, Haplotypes, ABC, ...) : devient possible pour des locus liés !

Remarque sur l'estimation à partir du déséquilibre de liaison

- ▶ Devient possible pour des locus liés !
- ▶ Mesure de la variation inter-locus (temps \Rightarrow espace)
- ▶ LD sur locus liés :
 - Locus proches: évènements anciens
 - Locus éloignés: évènements récents

Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size

Ben J. Hayes,^{1,4} Peter M. Visscher,² Helen C. McPartlan,¹ and Mike E. Goddard^{1,3}

¹Victorian Institute of Animal Science, Department of Natural Resources and Environment, Attwood, Victoria, 3049, Australia; ²University of Edinburgh, Edinburgh EH9 3JG, Scotland, UK; ³Institute of Land and Food Resources, University of Melbourne, Parkville, Victoria, 3052, Australia

Linkage disequilibrium (LD) between densely spaced, polymorphic genetic markers in humans and other species contains information about historical population size. Inferring past population size is of interest both from an evolutionary perspective (e.g., testing the “out of Africa” hypothesis of human evolution) and to improve models for mapping of disease and quantitative trait genes. We propose a novel multilocus measure of LD, the chromosome segment homozygosity (CSH). CSH is defined for a specific chromosome segment, up to the full length of the chromosome. In computer simulations CSH was generally less variable than the r^2 measure of LD, and variability of CSH decreased as the number of markers in the chromosome segment was increased. The essence and utility of our novel measure is that CSH over long distances reflects recent effective population size (N), whereas CSH over small distances reflects the effective size in the more distant past. We illustrate the utility of CSH by calculating CSH from human and dairy cattle SNP and microsatellite marker data, and predicting N at various times in the past for each species. Results indicated an exponentially increasing N in humans and a declining N in dairy cattle. CSH is a valuable statistic for inferring population histories from haplotype data, and has implications for mapping of disease loci.

The large number of densely spaced, polymorphic genetic markers generated by modern genomics is a powerful tool for answering genetic questions. For instance, they are being used to fine-scale-map trait genes (Pritchard and Przeworski et al. 2001) and to infer the history of the human population (Reich et al. 2001). Inferring past population size is of interest both from an evolutionary perspective (e.g., testing the “out of Africa” hypothesis of human evolution) and to improve models for the mapping of disease and quantitative trait genes.

Under a neutral model with constant effective population size (N), the homozygosity of a marker, the probability of sampling two identical alleles from the population, can be used to estimate N , provided the mutation rate is known (e.g., Kuhner et al. 1998; Slatkin and Bertorelle 2001). If N has changed in the past, the homozygosity will estimate a form of average N . The higher the mutation rate, the less sensitive to

Therefore, although estimates of average past population size from n unlinked loci can be more accurate than from n linked loci (Kuhner et al. 1998), using LD between n linked loci can provide additional information on historical changes in population size.

Most measures of LD, such as r^2 and related measures (Devlin and Risch 1995; Weir 1996), quantify the association between a pair of loci. Higher-order association coefficients analogous to r^2 can be defined for groups of 3, 4, or more loci, but they have not been found to be a practical value (Hill 1981). Such higher-order LD measures also ignore the essentially linear nature of chromosomes and of recombination. An ideal multilocus measure of LD would take account of this linearity and capture as much as possible of the information content of the data (no single statistic could contain all the information). In addition, it would be desirable if the measure of LD used by a commonly used statistic, at least on a standard

- 1 A. J. Drummond and A. Rambaut, “BEAST: Bayesian evolutionary analysis by sampling trees.,” BMC Evol. Biol., vol. 7, no. 1, p. 214, Jan. 2007.
- 2 A. J. Drummond, M. A. Suchard, D. Xie, and A. Rambaut, “Bayesian phylogenetics with BEAUti and the BEAST 1.7.,” Mol. Biol. Evol., vol. 29, no. 8, pp. 1969–73, Aug. 2012.

Doctorat ParisTech et Université d'Aarhus

THÈSE

pour obtenir le grade de docteur conjoint délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement (AgroParisTech)
et Université d'Aarhus**

Spécialité : Génétique animale

présentée et soutenue publiquement par

Belén JIMENEZ-MENA

RESEARCH

Analysis of different estimators of effective population size for conservation genetics

Belén Jiménez-Mena^{1,2,3*}, Etienne Verrier^{1,2} and Frédéric Hospital^{1,2}

Abstract

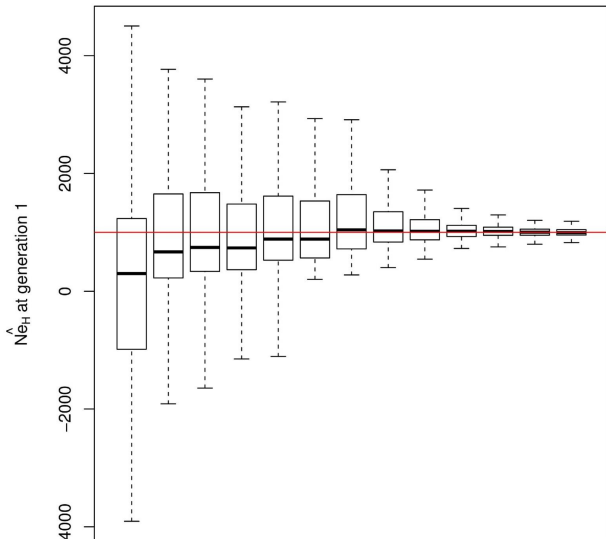
Background: The difficulty in obtaining accurate estimates of the effective population size (N_e) has been remarked across the literature. There are different methods to estimate N_e that use different kind of information about pedigree or molecular markers to measure genetic drift. The analysis and comparison of the performance of methods that infer N_e using molecular information are gaining importance due to the decreasing costs of sequencing and the availability of large number of molecular markers. In this simulation analysis, we (i) analysed in detail an estimator based on the decrease of heterozygosity, \hat{N}_{eH} , and (ii) performed a comparison of this with an estimator based on linkage disequilibrium, \hat{N}_{eLD} , and standardized variance in allele frequencies, \hat{N}_{eT} , using different scenarios of number of loci, population sizes, initial allele frequencies and time interval between samples.

Results: \hat{N}_{eH} presented bias and a large variability over time, explained by the large variability of heterozygosity and its asymmetrical distribution. We showed that the combination of the number of segregating alleles and the number of loci used might be provoking an impact on the variability of \hat{N}_{eH} . When using low number of loci (<100), the three estimators displayed differences in their performance after one generation of genetic drift. With higher number of loci (≥ 100), the mean of the three estimators narrowed towards the true value of N_e and their performance was improved. \hat{N}_{eT} and \hat{N}_{eLD} were the best estimators to use in such case. All estimators performed accurately when initial allele frequencies were close to 0.5; \hat{N}_{eT} seemed to have the best performance when the initial allele frequencies were increased.

Conclusions: \hat{N}_{eT} and \hat{N}_{eLD} estimators are the most adequate to use under the conditions tested in this study. \hat{N}_{eH} might be more efficient with a large number of loci with high number of alleles. To have an accurate estimate of N_e , a minimum number of 100 loci should be used. When the loci used have allele frequencies different than 0.5, attention on potential bias on the estimation on N_e should be considered.

Keywords: effective population size; genetic drift; heterozygosity; temporal method; linkage disequilibrium method

Figure 3: N_{eH} values obtained at generation $t = 1$ against the number of alleles. The black and red lines represent, respectively, the median and the true value of N_e . Outliers were not included in the figure.



Submitted as a research article for BIOLOGY LETTERS.

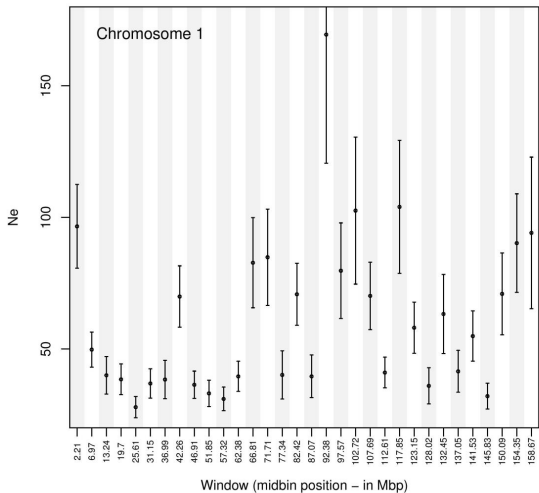
Word count: 2758

One size fits all?

Direct evidence for the heterogeneity of genetic drift throughout the genome.

Belén Jiménez-Mena^{1,2,3}, Paula Tataru¹, Rasmus F Brøndum⁴, Goutam Sahana⁴, Bernt GuldbRANDtsen⁴, Thomas Bataillon¹.

1. Bioinformatics Research Center (BiRC), Aarhus University, Aarhus Denmark.
2. INRA, UMR1313 Génétique animale et biologie intégrative, 78350 Jouy-en-Josas, France
3. AgroParisTech, UMR1313 Génétique animale et biologie intégrative, 16 rue Claude Bernard, 75231 Paris 05, France
4. Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark



1B. Example of within-chromosome heterogeneity in estimated N_e . Each dot