

Introduction to effective population size and demographic inference

Brendan Reid and Marianne Dehasque
July 9th, 2025
UP Cebu, Cebu City, Philippines

Overview

- Introduction: what is demography, and what is N_e ?
- Effective size vs. census size
- N_e and genetic diversity loss
- General methods for estimating N_e
- N_e over different time scales

Demography = the study of populations

Why are we interested in population size?

Demography = the study of populations

Population size is important because it affects...

- extinction risk
- sustainable yield / stock assessments
- efficacy of selection
- rates of genetic drift and inbreeding

Demography is hard!

Many populations are difficult to directly measure

- Individuals are hard to count
- Reproduction, mortality, migration can be difficult to directly observe
- Our observation window is often very short

Genetics can provide us an indirect window into understanding demography

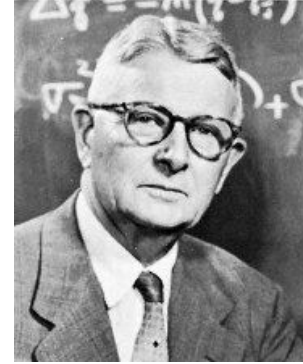
Genetics and population size: history

EVOLUTION IN MENDELIAN POPULATIONS

SEWALL WRIGHT

University of Chicago, Chicago, Illinois

Received January 20, 1930



THE
GENETICAL THEORY OF
NATURAL SELECTION

BY
R. A. FISHER, Sc.D., F.R.S.

Wright-Fisher population models

Standard models for population genetics

In their simplest form:

- No migration
- Random mating
- Constant **effective population size**

Effective population size is a key parameter!

What the N_e is effective population size (N_e)?

The population number

It will be well to discuss more fully, before going on, what is to be understood by the symbol N used here for population number. The conception is that of two random samples of gametes, N sperms and N eggs, drawn from the total gametes produced by the generation in question ($N/2$ males and $N/2$ females each with a double representation from each series of allelomorphs). Obviously N applies only to the breeding population and not to the total number of individuals of all ages. If the population fluctuates greatly, the effective N is much closer to the minimum number than to the maximum number. If there is a great difference between the number of mature males and females, it is closer to the smaller number than to the larger. In fact, as just shown, a population of N_m males and an indefinitely large number of females is approximately equivalent to a population of $4N_m$ individuals equally divided between males and females.

The conditions of random sampling of gametes will seldom be closely approached. The number of surviving offspring left by different parents may vary tremendously either through selection or merely accidental causes, a condition which tends to reduce the effective N far below the actual number of parents or even of grandparents. How small the effective N of a population may be is indicated by recent studies of SMITH and CALDER (1927) on the Clydesdale breed of horses in Scotland, in which they find a steady increase in the degree of inbreeding (Coefficient F) equivalent to that in population headed by only about a dozen stallions. Even more striking is the rapid increase in the coefficient of inbreeding in the early history of the Shorthorn breed of cattle (McPHEE and WRIGHT 1925).

Related to the number of individuals in a population that actually contribute genes to the next generation

Influences how quickly gene frequencies will change due to random chance and inbreeding will occur in a population

What is effective population size **not**?

- Different from census population size (N_c)
- N_e is usually $< N_c$
 - Unequal sex ratios
 - Unequal reproductive success across individuals

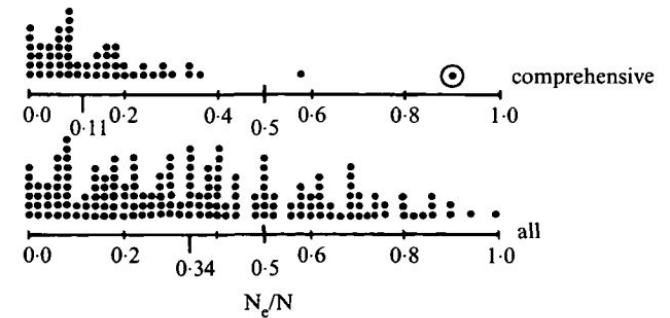


Fig. 1. Distributions of estimates of effective/actual population size (N_e/N) ratios. Comprehensive estimates (that include the effects of fluctuation in population size, variance in family size and unequal sex-ratio) are above and all estimates below. The circled outlier is for a pair mated rainbow trout (*Oncorhynchus mykiss*) population. Means of estimates are indicated below vertical lines.

Marine populations and N_e

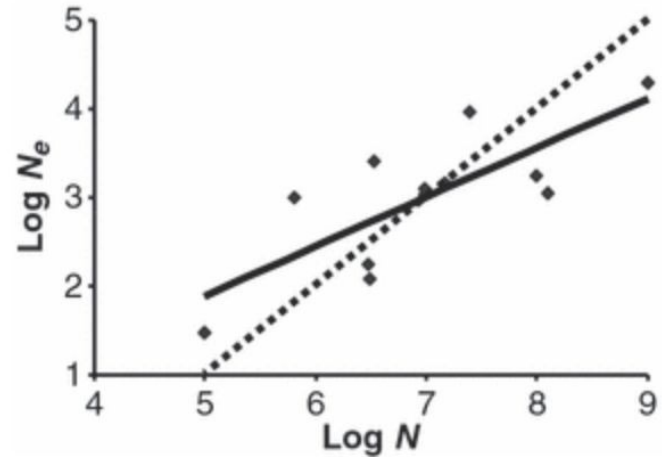
Given what we know about marine populations, would you expect N_e for marine populations would be generally smaller or larger than N_e for terrestrial or freshwater populations?

Marine populations and N_e

Marine census population sizes can be much larger than populations for species of other types...

However, N_e is often 2-6 orders of magnitude lower!

Why would this be?

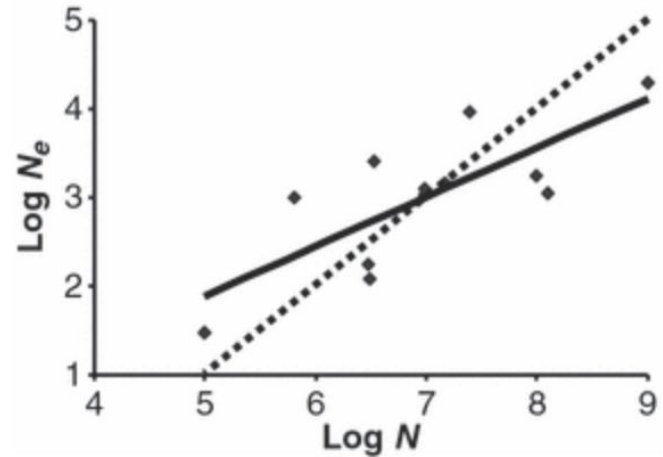


Hauser + Carvalho, 2008

Marine populations and N_e

High variance in reproductive success among individuals ('sweepstakes reproduction') for many marine species

Spatial variation in reproductive success



Hauser + Carvalho, 2008

Is N_e still relevant quantity to measure

Yes!

- *Informative about evolutionary processes
- *Linked to census size



But proceed with caution...

- *Interpretation depends on understanding N_e/N_c
- *Many factors can confound our estimates (migration, population structure, etc)



Genetic diversity loss in the Anthropocene

MOISES EXPOSITO-ALONSO , TOM R. BOOKER , LUCAS CZECH , LAUREN GILLESPIE , SHANNON HATELEY , CHRISTOPHER C. KYRIAZIS, PATRICIA L. M. LANG

, LAURA LEVENTHAL , DAVID NOGUES-BRAVO , [...], AND ERIN ZESS 

+5 authors

[Authors Info & Affiliations](#)

MOLECULAR ECOLOGY


From the Cover |  **Full Access**

Meta-analysis reveals lower genetic diversity in overfished populations

Malin L. Pinsky  Stephen R. Palumbi

First published: 11 September 2013 | <https://doi.org/10.1111/mec.12509> | Citations: 285

Time-lagged genomic erosion and future environmental risks in a bird on the brink of extinction

Xufen Liu, Ester Milesi, Claudia Fontserè, Hannah L. Owens, Robert Heinsohn, M. Thomas P. Gilbert, Ross Crates, David Nogués-Bravo and Hernán E. Morales 

Published: 26 March 2025 | <https://doi.org/10.1098/rspb.2024.2480>

REVIEWS AND SYNTHESSES

 **Open Access**



Estimated six per cent loss of genetic variation in wild populations since the industrial revolution

Deborah M. Leigh  Andrew P. Hendry, Ella Vázquez-Domínguez, Vicki L. Friesen

First published: 07 May 2019 | <https://doi.org/10.1111/eva.12810> | Citations: 168

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

 **Open Access**



Temporal Loss of Genome-Wide and Immunogenetic Diversity in a Near-Extinct Parrot

Luke W. Silver, Katherine A. Farquharson, Emma Peel, M. Thomas P. Gilbert, Katherine Belov, Hernán E. Morales, Carolyn J. Hogg 

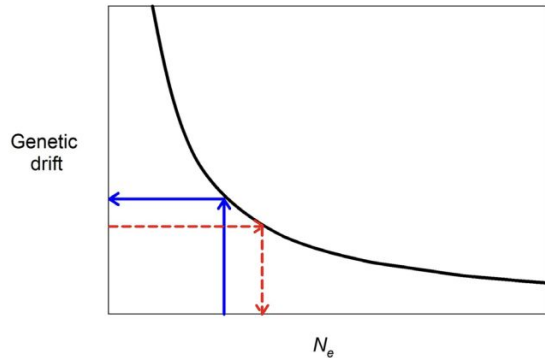
First published: 25 March 2025 | <https://doi.org/10.1111/mec.17746>

Why is N_e useful in the context of genetic diversity loss?

- Allows us to quantify and predict one component of population genetic change (drift)
 - Waples 2025: “ N_e provides a handy way to quantify the random effects of genetic drift and resulting consequences for rates of inbreeding and loss of genetic diversity.”
- Provides guides for conservation based on N_e values needed to minimize inbreeding and preserve adaptive genetic diversity over the long term (50/500 rules)

How do we go from N_e to rates of drift?

- In an *ideal** population, an allele with frequency p will become **fixed** in (on average) $4N_e[p \ln(p) + (1-p) \ln(1-p)]$ generations
- So, larger N_e = slow genetic drift, smaller N_e = fast genetic drift



*AKA: 'Wright-Fisher' population

*randomly mating

*constant size

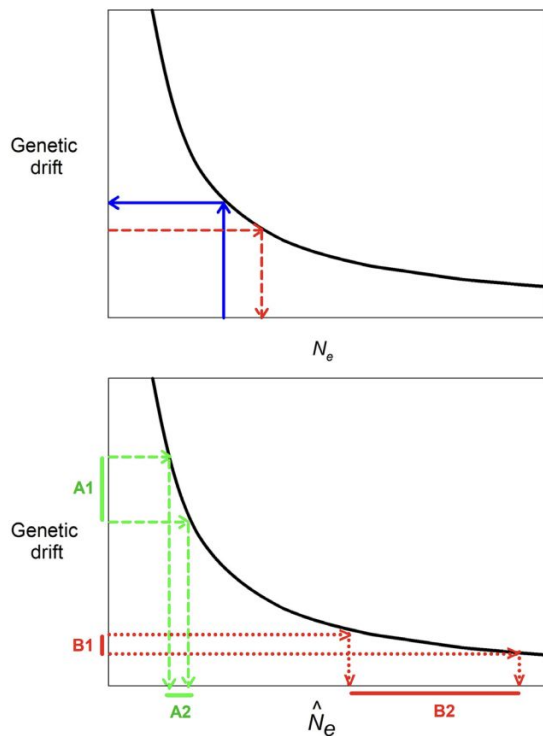


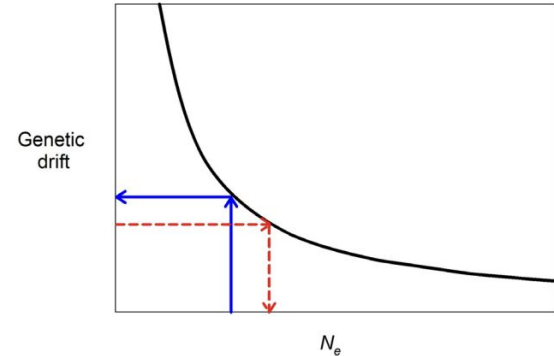
FIGURE 1 | Top: The non-linear, inverse relationship between N_e and genetic drift (black curve). Knowing N_e allows one to predict the rate of genetic drift (solid blue arrows). At some point, researchers realised that if they could measure an index of genetic drift, they could reverse the process and use that information to estimate N_e (dashed arrows). Bottom: Obtaining robust estimates is easiest for small populations (dashed green lines), for which the genetic drift signal is strong and substantial uncertainty in measuring the drift index (A1) has relatively little effect on \hat{N}_e (A2). The situation is reversed in large populations (dotted red lines): Even a small error in estimating the genetic index (B1) has a large effect on \hat{N}_e (B2).

Generally, more uncertainty in estimating N_e for large populations...

However, smaller errors in estimating small N_e can still result in large errors in predicting drift (because $N_e \sim$ drift relationship is non-linear)

How can we estimate N_e with genetic data ?

- Temporal change in allele frequency
 - Essentially, measure genetic drift over time and fit a Wright-Fisher model
 - Estimate corresponds to geometric average of N_e over time
 - Implemented in **NeEstimator**



How can we estimate N_e with genetic data ?

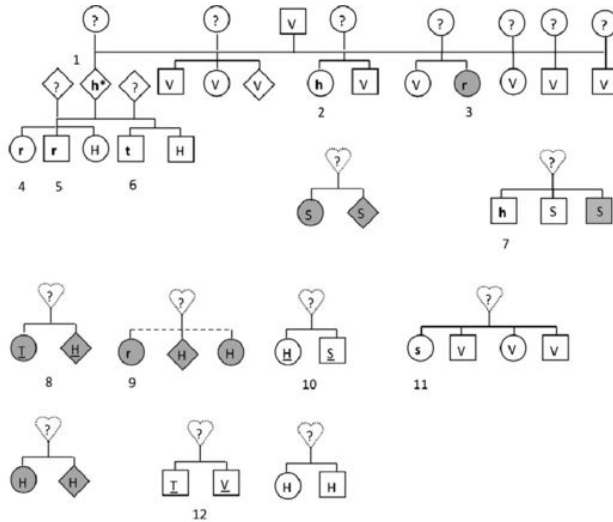
- Linkage disequilibrium (LD)
 - Contrast with linkage equilibrium - random inheritance of alleles on different chromosomes
 - We expect more non-random associations of alleles (= higher LD) in small populations than in large populations
 - Some methods provide estimates of contemporary N_e (**NeEstimator**)
 - Some methods can incorporate information from physically linked alleles to estimate change over time (**GONE**)

How can we estimate N_e with genetic data ?

- Model-based methods
 - Fit allele frequencies to population models
 - Various methods (**momi2**, dadi, stairwayplot, etc.)

How can we estimate N_e with genetic data ?

- Kinship/pedigree
 - Use distribution of relatives to estimate N_e



E.g.:

COLONY (Jones and Wang 2009)
close-kin mark-recapture (CKMR;
Ruzzante et al. 2019)

How can we estimate N_e with genetic data ?

- Coalescent-based methods
 - Based on distribution of coalescence times in the genome

LETTER

doi:10.1038/nature10291

Inference of human population history from individual whole-genome sequences

Heng Li^{1,2} & Richard Durbin¹

Received: 28 May 2019 | Revised: 11 October 2019 | Accepted: 12 November 2019
DOI: 10.1002/evl.35888

REVIEW ARTICLE

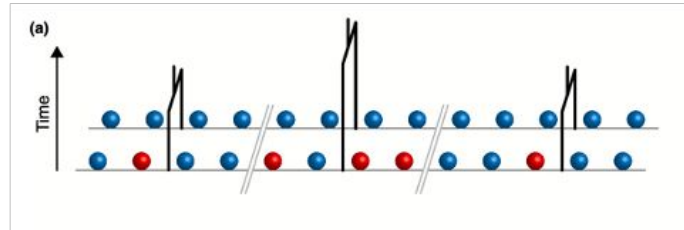
Ecology and Evolution

WILEY

A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data

Niklas Mather | Samuel M. Traves | Simon Y. W. Ho

PSMC input data: genome sequence from a single diploid individual



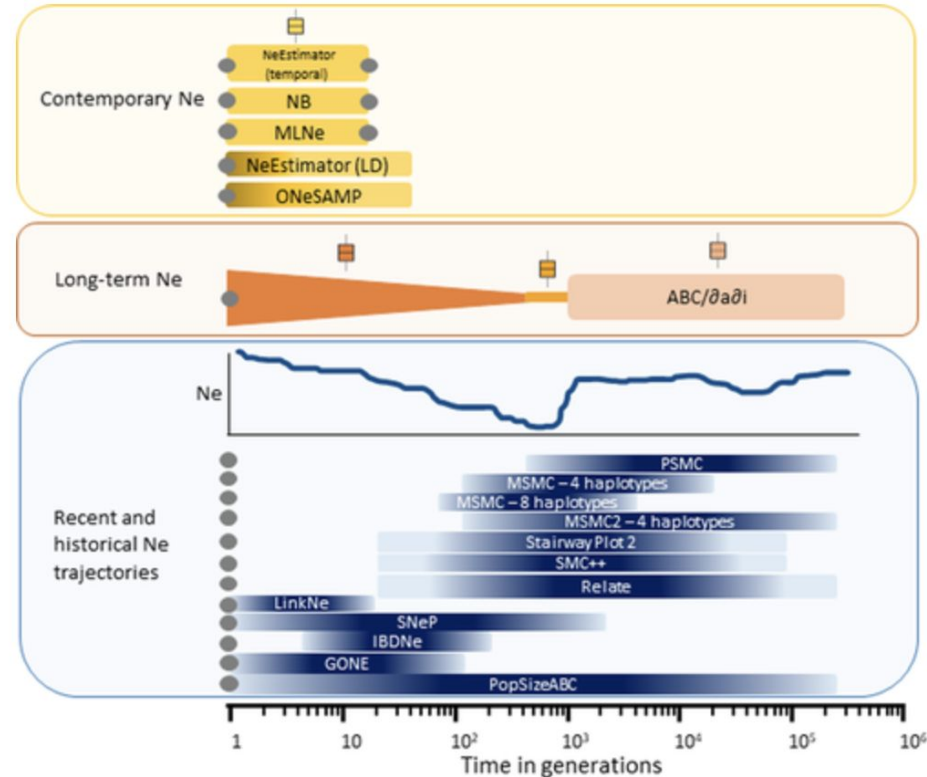
Also MSMC, SMC++,
Bayesian Skyline
Plots, etc...

How do we estimate N_e with genetic data ?



Choosing a method

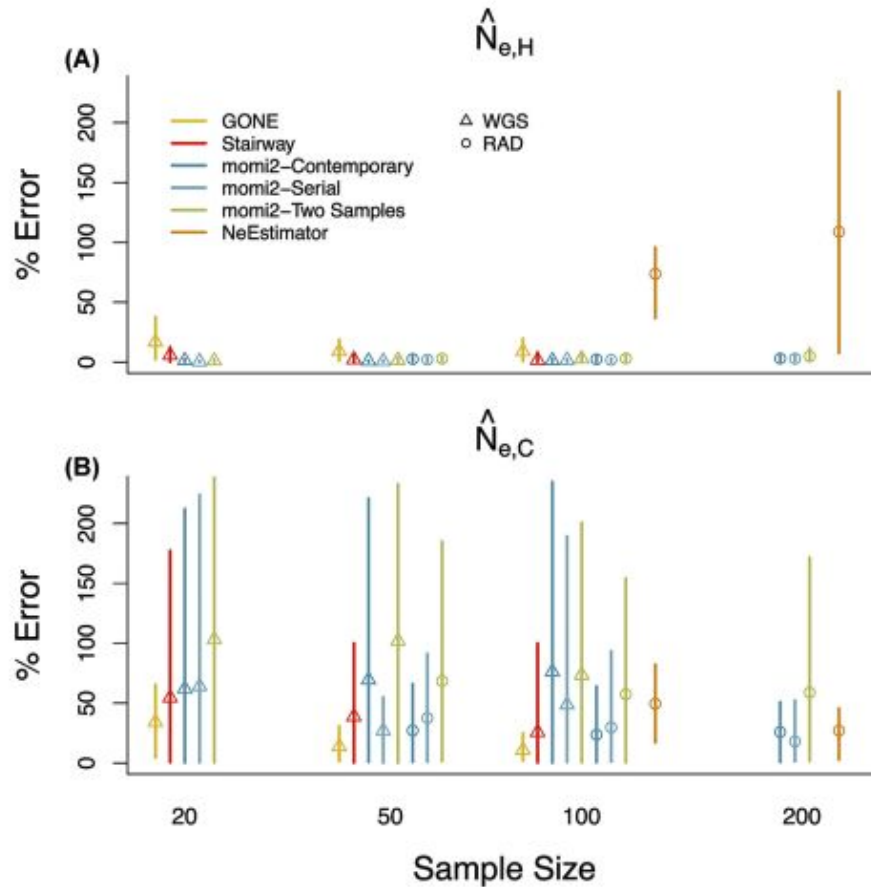
What time scale are you interested in?



Choosing a method

What kind of data do you have?

- *Data type (microsats, SNPs, WGS, Sanger sequencing)
- *Number of individuals
- *Number of loci
- *Temporal sampling scheme



Historical N_e can be easier to accurately measure than contemporary N_e

Fig. 3. Mean absolute percentage error for estimating (A) historic effective population size ($N_{e,H}$) and (B) contemporary effective population size ($N_{e,C}$). Error bars show 10%–90% quantiles for each.

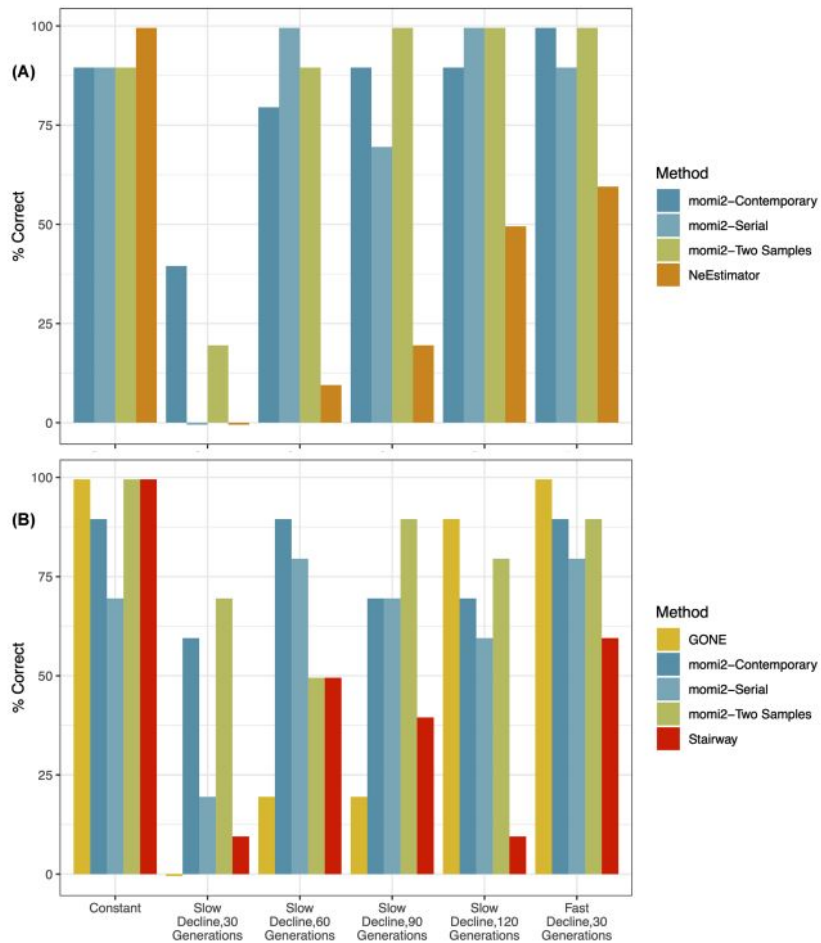


Fig. 2. Power for detecting the correct demographic scenario for (A) RAD data and (B) WGS data. Results are shown for the largest sample size for each data type (RAD $n = 200$, WGS $n = 100$).

Power to detect
demographic declines
depends on method,
genomic data type, and
the rate of the decline.

Questions?

Get familiar with the platform (GitHub)