A century-long genetic record reveals that protist effective population sizes are comparable to those of macroscopic species

Phillip C. Watts1,2, Nina Lundholm3, Sofia Ribeiro4 and Marianne Ellegaard5,†

1Institute of Integrative Biology, University of Liverpool, Biosciences Building, Liverpool L69 7ZB, UK
2Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, FIN-40014, Jyväskylä 40014, Finland
3The Natural History Museum of Denmark, Sølvgade 835, DK-1307 Copenhagen, Denmark
4Department of Marine Geology and Glaciology, Geological Survey of Denmark and Greenland, Øster Voldgade 10, DK-1350 Copenhagen, Denmark
5Department of Biology, University of Copenhagen, Øster Farimagsgade 2D, DK-1353 Copenhagen, Denmark

Effective population size ($N_e$) determines the rate of genetic drift and the relative influence of selection over random genetic changes. While free-living protist populations characteristically consist of huge numbers of cells ($N$), the absence of any estimates of contemporary $N_e$ raises the question whether protist effective population sizes are comparably large. Using microsatellite genotype data of strains derived from revived cysts of the marine dinoflagellate Pentapharsodinium dalei from sections of a sediment record that spanned some 100 years, we present the first estimates of contemporary $N_e$ for a local population in a free-living protist. The estimates of $N_e$ are relatively small, of the order of a few 100 individuals, and thus are similar in magnitude to values of $N_e$ reported for multicellular animals: the implications are that $N_e$ of $P$. dalei is of many orders of magnitude lower than the number of cells present ($N_e/N \approx 10^{-12}$) and that stochastic genetic processes may be more prevalent in protist populations than previously anticipated.

1. Introduction

A crucial concept in evolutionary biology is that only some individuals obtain genetic representation in subsequent generations. This number of ‘successful’ individuals may be approximated by the contemporary effective population size ($N_e$), which, strictly defined, is the size of an ‘ideal’ genetic population experiencing the same amount of random genetic changes as the target population [1,2].

Our use of contemporary $N_e$ refers to a local population and therefore is different to the ‘long-term $N_e$’, that represents an expected (given some estimate of mutation rate) amount of neutral genetic variation maintained within a species at mutation-drift equilibrium [2]. Estimates of $N_e$ predict the rate of loss of genetic diversity and the efficacy of selection, with the action of selection compared with random genetic drift favoured in large effective populations [1,2]. Further insights can be gained from the ratio $N_e/N$, where $N$ is the number of potentially reproductively active individuals, which quantifies the extent of departure from idealized population genetic models and the susceptibility of apparently large populations to stochastic evolutionary processes [3–5].

The potentially huge population sizes of free-living, unicellular eukaryotes (protists) would appear to render their populations somewhat unsusceptible to random genetic changes. However, we are not aware of any estimates of contemporary $N_e$ in any free-living protist. This lack of data raises fundamental questions about the magnitude of contemporary protist effective population sizes.
sizes as well as the extent to which estimates of $N_e$ are less than the absolute numbers of cells present.

Obtaining reasonable estimates of contemporary $N_e$ in field populations is notoriously difficult [1–6]. The most widely used genetic methods to estimate $N_e$ exploit the association between population size and the magnitude of (stochastic) temporal changes in allele frequencies [1,6–8]. Such temporal estimates of $N_e$ require genotypes from two or more samples separated by several generations, which are not available for most taxa unless there have been sufficient foresight and resources to establish a long-term sampling programme and maintain appropriate material archives.

Obtaining temporally separated samples can be straightforward for many marine protists, such as numerous dinoflagellates and diatoms, when they produce resting stages (cysts) that accumulate in benthic sediments. Proist cysts may remain viable for up to 100 years [9,10], thereby providing a source of naturally archived historic material. We revived cysts of *Pentapharsodinium dalei*, a spring-blooming marine dinoflagellate that inhabits polar, sub-polar and cold-temperate coastal regions, from samples spanning some 100 years to make the first estimates of contemporary $N_e$ in a free-living protist population.

### 2. Material and methods

All viable *P. dalei* cysts were recovered and cultured from four slices of sediment core taken from Koljö Fjord, Sweden (58°13’ N, 11°34’ E). Estimated dates for the core slices were 2006, 1985 (±3 years s.e.), 1960 (±5 years s.e.) and 1922 (±12 years s.e.) [9,10], representing changes in population composition over approximately a century (84 ± 12 years s.e.); these slices yielded 41, 31, 33 and 29 clonal strains, respectively, with each strain representing a single haploid cell isolated from a cell culture of a germinated cyst. Observations indicate that the cysts of *P. dalei* are hypnozygotes (diploid cells resulting from sexual fusion), and thus part of the sexual cycle [11]. The frequency of sexual reproduction has not been determined for this species, but as cysts are considered to be survival stages that are produced most frequently at the end of bloom periods, it is a reasonable assumption that *P. dalei* reproduces sexually once a year.

Genomic DNA from monoclonal cultures was extracted using a CTAB method [11]. Samples were genotyped at six microsatellite loci [11]. PCRs consisted of 0.3 μM forward and reverse primers (forward primers 5'-labelled with 6-FAM, NED or HEX fluorophores; Applied Biosystems), 10 μl HotStarTaq Master Mix (Qiagen) and published thermal cycling conditions [11]. PCR products were pooled with Genescan LIZ500, separated by capillary electrophoresis on an ABI3730xl and sized using GENEMAPPER (Applied Biosystems).

Contemporary estimates of $N_e$ (and 95% CI) for all sample pairs were made using two temporal methods. First, MLNE was used to estimate $N_e$ using a maximum-likelihood (ML) estimator [8] that does not assume a particular mating system or require genetic equilibrium, and applies to small and large populations. Second, we used NEESTIMATOR v. 2 [12] to calculate $N_e$ using Waples’ moment method [6] based on Nei & Tajima’s [13] estimator of standardized variance in allele frequency change. Our calculations assumed a 1-year generation time (the timing of the sexual cycle), a closed population (Koljö Fjord has limited water exchange with elsewhere (see electronic supplementary material)) and a population not at migration–drift equilibrium.

The total abundance of cells ($N$) in Koljö Fjord was estimated using data on the densities of *P. dalei* in the neighbouring Gullmar Fjord and the bathymetry of Koljö Fjord (see electronic supplementary material) and are estimates of the numbers of the vegetative (increasing by cell division) stages at a given time.

### 3. Results

All loci were polymorphic, yielding a total of 80 alleles (see electronic supplementary material). Genotyping revealed some haplotypes originating from the same cyst, which left 21, 27, 20 and 18 different haploid genotypes for analysis. One locus failed to amplify alleles in one to two clones per sample. There was no significant difference among samples in the frequency of non-amplifying genotypes (Kruskal–Wallis $\chi^2 = 0.602$, d.f. = 3, $p = 0.896$) and the presence of null alleles would have little impact upon estimates of $N_e$ [8,14]; treating missing data as distinct (null) alleles does not affect the basic results (see electronic supplementary material).

None of our estimates of $N_e$ in *P. dalei* was particularly huge. The ML estimates of contemporary $N_e$ of *P. dalei* among pairs of samples in Koljö Fjord lay between 179 and 815, with a maximum upper 95% CI of 2489; comparable, albeit slightly higher, estimates of $N_e$ were derived using the moment estimator ($N_e = 178–1183$, with a maximum upper 95% CI = 5545; table 1). A moderate $N_e$ remained when we examined the potential effects of error associated with sediment dating (see electronic supplementary material).

Based on a conservative value of 1500 cells $m^{-1}$ in the water column above the halocline, the number of vegetative cells of *P. dalei* in Koljö Fjord was estimated as approximately $2.9 \times 10^4$ (see electronic supplementary material). While these calculations are approximate, the $N_e$ for *P. dalei* represents a tiny fraction of the number of cells present ($N$), with an $N_e/N \sim 10^{-12}$ (and between approx. $10^{-11}$ and approx. $10^{-13}$ based on the maximum and minimum 95% CIs).

### 4. Discussion

Palstra & Ruzzante’s [4] meta-analysis yielded an average contemporary $N_e \sim 260$ for wild populations. Our first estimates of contemporary $N_e$ in a free-living protist population are robust to error associated with sediment dating and are mostly of the same order of magnitude as those of macroscopic species.
A corollary of the numerical abundance of protist cells is an extremely low $N_e/N$ ratio of approximately $10^{-12}$ in *P. dalei*, which is the lowest value for any taxon that we are aware of. For comparison, median estimates of $N_e/N$ for animals are around 0.1–0.14 [3,4], but this ratio typically is lower (approx. $10^{-3}$ to $10^{-5}$) in marine teleosts [4,5,15,16] and approaches approximately $10^{-8}$ in freshwater copepods [14]; the single estimate of $N_e/N$ in an abundant marine macroalga, *Fucus serratus*, was between $10^{-3}$ and $10^{-4}$ [17].

The ecological and life-history processes behind *P. dalei’s* low $N_e/N$ are not known but are probably a consequence of temporal fluctuations in population size, as well as variation among strains in achieving genetic representation [3] and extensive vegetative cell division (that increases $N$). Also, our study population may have violated some of the model requirements. For example, Koljø Fjord is not completely closed and estimates of $N_e$ based on an assumption of no gene flow can be biased compared with estimates of $N_e$ that account for gene flow [8,18], with the direction and magnitude of the bias determined by the genetic composition of the source population(s) [4,14] and the amount of gene flow [18]; for example, under very high gene flow (i.e. where all individuals within a local population, for example in Koljø Fjord, are derived from the global population), temporal estimates of $N_e$ for a local population are expected to be biased downwards by up to 50% owing to an additional variance component [18]. Future studies of protists should consider this effect by quantifying spatial genetic structure and by jointly estimating $N_e$ and rates of migration [8]. The impact upon our estimates of $N_e$ of other aspects of protist life history, such as a lack of separate sexes or overlapping generations owing to possible revival of buried cysts, is not clear. However, the long interval between slices of sediment core should provide temporal estimates of $N_e$ with increased precision and reduced bias (owing to any effect of overlapping generations) [5]. Low $N_e/N$ ratio may reflect the action of selection, with hitchhiking effects expected to be important when there is substantial clonal reproduction [2], and insufficient sampling is expected to be an issue when $N_e$ is very large [4,5]. Irrespective of the underlying driver(s), an extremely low $N_e/N$ ratio appears to be characteristic of numerically abundant multi-cellular aquatic organisms [4,5,14–17] and aquatic protists apparently are no exception. The implication of low contemporary $N_e$ is that stochastic genetic processes are more prevalent than anticipated from a population census.

One key issue with the concept and interpretation of $N_e$ in wild populations is that it may be defined in many ways [1,2,5], including a long-term $N_e$ (defined in the Introduction). Although estimates of long-term $N_e$ are perhaps less relevant to understanding contemporary evolutionary processes [5], the few data for protists provide evidence for comparatively large [19,20] and small [21] long-term $N_e$. Given the diversity of protist life histories, we expect that additional studies on protists will find comparable inter- and intra-specific variation in contemporary $N_e$: understanding the drivers behind any diversity in $N_e$ presents an exciting challenge. With this in mind, contemporary $N_e$ can be estimated from a single genetic sample [1,22], providing the advantage that every sample gives an estimate of $N_e$ [22]; moreover, some single sample estimators of $N_e$ outperform temporal estimators of $N_e$ when population sizes are small and there are large sample sizes [22]. By contrast, temporal estimators of $N_e$ are appropriate for characterizing large populations, when there is a long time interval between samples [22], as here, and when there is high gene flow [18]. We did not use single sample estimators of $N_e$ because (i) they suffer from poor statistical power when few loci are used [22] and (ii) an analysis of haploid data represented an important departure from their underlying model assumptions. Nonetheless, for protists that can be genotyped at many loci and during a diploid phase, single sample estimators of $N_e$ could also be used [22]. For most protists, however, there remains an apparent need for temporally separated samples to obtain accurate estimates of $N_e$, and this should not present a major obstacle for species that produce cysts as they are amenable to the palaeogenetic approach used here (see also [9,10,23]).

Acknowledgements. We would like to thank Jinliang Wang, David Tallmon and Robin Waples for their helpful comments about estimating $N_e$ with haploid samples, and also Adam Eyre-Walker, Daniel Ruzzante and the editors for their insightful critiques of the manuscript. Data accessibility. Genotypes archived with Dryad: doi:10.5061/dryad.22116.

Funding statement. The authors are part of PRODIVERSA, a NordForsk-funded network on protist population genetics. Laboratory work was financially supported by the Danish Research Council (2111-04-0011) and fieldwork by Göteborg University Marine Research Centre. S.R. holds a postdoctoral fellowship from the Carlsberg Foundation (2011_01_0337).

References


