MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3246–3249

NEWS AND VIEWS

COMMENT

Heterozygosity of the Yellowstone wolves

IVANA JANKOVIC,* BRIDGETT M. vonHOLDT† and NOAH A. ROSENBERG*‡

*Center for Computational Medicine and Bioinformatics, University of Michigan, 2017 Palmer Commons, 100 Washtenaw Ave., Ann Arbor, MI 48109, USA, †University of California, Los Angeles, Ecology & Evolutionary Biology, 621 Charles E. Young Dr. South, Los Angeles, CA 90095, USA, ‡Department of Human Genetics and the Life Sciences Institute, University of Michigan, 2017 Palmer Commons, 100 Washtenaw Ave., Ann Arbor, MI 48109, USA

Received 29 December 2009; revision received 26 April 2010; accepted 12 May 2010

In a recent study, vonHoldt *et al.* (2008) examined the success of the grey wolf (*Canis lupus*) re-introduction programme into Yellowstone National Park in preserving the genetic variation of the population. They evaluated a variety of aspects of genetic diversity in the wolf population, which originated from 41 founders introduced in 1995 and 1996 and which has remained genetically isolated since the reintroduction. In each of a large number of individuals sampled during the initial recovery period, 1995–2004, vonHoldt *et al.* (2008) genotyped 26 microsatellite loci. Their analyses, which included estimates of mean observed and expected heterozygosity (H_O and H_E, respectively), generally indicated that this isolated wolf population is effective at inbreeding avoidance and maintenance of genetic diversity. However, some aspects of their genetic variation analyses appeared to be somewhat incompatible. Levels of expected heterozygosity, calculated using Nei's (1987) heterozygosity estimator ($\hat{H}_E$), identified a decreasing trend in genetic variation starting from 1997, after the introductions were complete (Fig. 1a). The authors suggested that if this trend continues, wolf fitness might decrease through the negative effects of inbreeding and reduced adaptability. Curiously, the reported $\hat{H}_O$ showed the opposite trend to $\hat{H}_E$ (Fig. 1a), demonstrating increasing proportions of heterozygous individuals, potentially indicative of a reduction in inbreeding over time. $\hat{H}_O$ was consistently lower than $\hat{H}_E$, however, a result that might be suggestive of inbreeding. As behavioural observations documented very few cases of inbreeding over the 10 years of the study (vonHoldt *et al.* 2008), it is likely that factors other than inbreeding have contributed to the discrepancy between $\hat{H}_E$ and $\hat{H}_O$.

Correspondence: Ivana Jankovic, Fax: (734) 615 6553; E-mail: ivanaj@umich.edu

The Yellowstone wolf data set of vonHoldt *et al.* (2008) was unusually enriched for close relatives, because of the small size of the founding population ancestral to all sampled individuals, the lack of gene flow from outside immigrants, the mating hierarchy and high variance of reproductive success in the species and the near-comprehensive sampling of the population (considering annual census sizes, the per-year proportion of the population sampled was as high as ~86%). Recent developments in the estimation of allele frequencies from inbred and related samples (e.g. Weir 1996; Broman 2001; Bourgain *et al.* 2004; DeGiorgio & Rosenberg 2009) have demonstrated that the presence of close relatives in a sample introduces a downward bias in $\hat{H}_E$, providing a possible explanation for the unusual heterozygosity observations of vonHoldt *et al.* (2008). We were, therefore, interested in determining whether accounting for the bias in $\hat{H}_E$ caused by the inclusion of relatives would affect the conclusions of vonHoldt *et al.* (2008) regarding temporal trends in wolf genetic variation.

A newly developed unbiased estimator for heterozygosity ($\tilde{H}_E$) accounts for the presence of close relatives when kinship coefficients (Φ) between individuals in the sample are known (DeGiorgio & Rosenberg 2009). We applied $\tilde{H}_E$ to genotype and kinship data for the wolves, separately analysing data from each of the 10 years of the study. Data were taken from vonHoldt *et al.* (2008), employing a pedigree that had previously been constructed using a combination of field observations and pairwise allele sharing. To adjust for levels of relatedness in the computation of $\tilde{H}_E$, for each year, at each locus, we first calculated the average pairwise kinship coefficient ($\bar{\Phi}$) across pairs of individuals sampled at the locus (Fig. 2). To determine Φ between pairs of wolves, we used inferred relationships from wolf pedigrees and the algorithm of Lange (2002, pp. 81–83), as implemented by Atkinson & Therneau (2008). For individuals with two unknown parents, we considered the unknown parents to be founders unrelated to all sampled individuals. In rare instances in which the identity of only one parent was uncertain, we considered possible half-siblings to be full-siblings. In computing both heterozygosity and $\bar{\Phi}$ at a locus, we excluded from calculations at that locus individuals for which data were missing. Calculations applied to samples with missing data excluded in this manner are indicated by a 'prime' (e.g. $\tilde{H}'_E$). After estimating per-locus heterozygosities, we averaged them across loci to obtain overall annual estimates.

When the downward bias introduced by the inclusion of relatives is taken into account through the use of kinship coefficients, in the period after the introductions, the mean

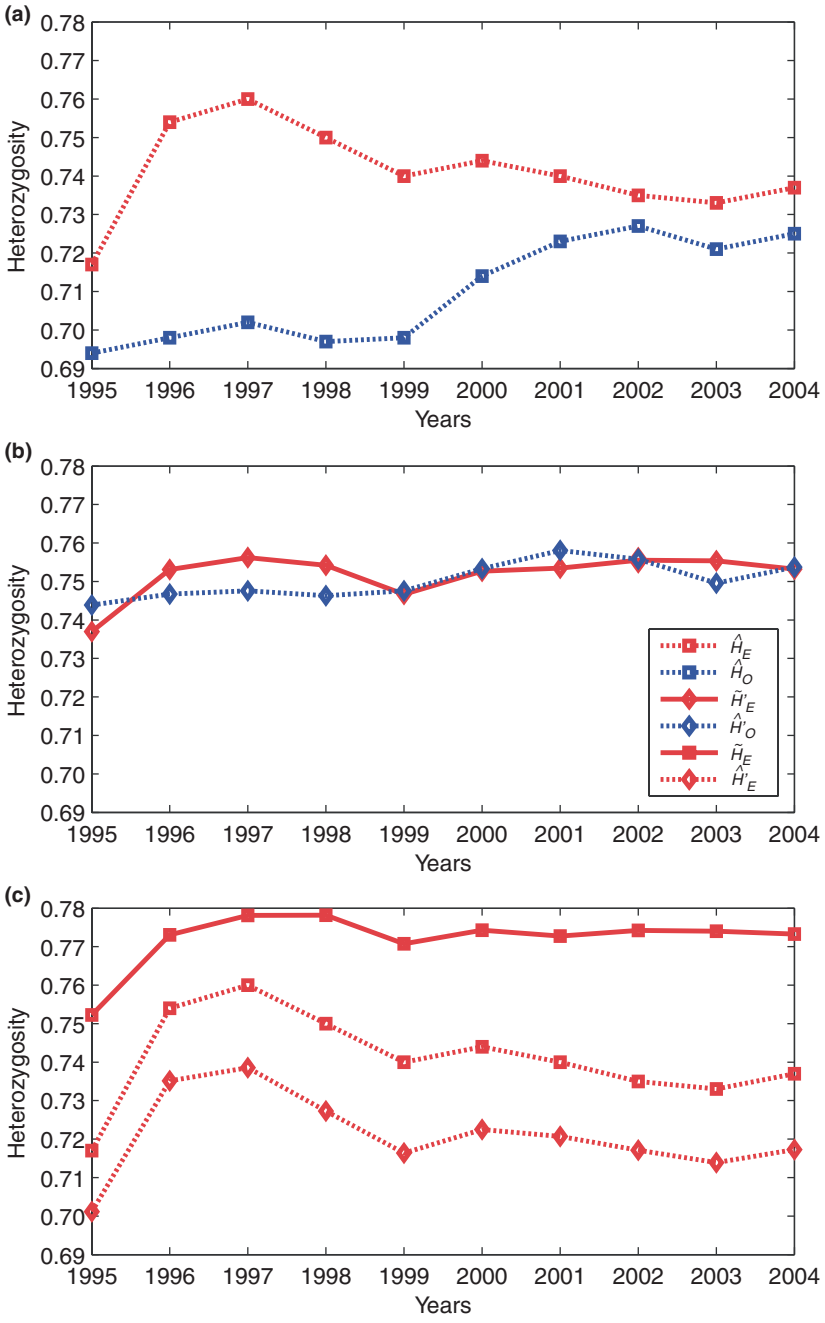

Fig. 1 (a) Annual values of $\hat{H}_E$ and $\hat{H}_O$ from Table 1 in vonHoldt *et al.* (2008). (b) Annual values of $\tilde{H}'_E$ from mean locus heterozygosities calculated using the DeGiorgio & Rosenberg (2009) estimator and excluding missing data. The corresponding annual values of $\hat{H}'_O$ calculated by excluding missing data are presented for comparison. (c) Annual values of $\tilde{H}_E$, $\hat{H}_E$ and $\hat{H}'_E$ averaged across all loci. $\tilde{H}_E$ and $\hat{H}_E$ treat missing data as an additional allele, whereas $\hat{H}'_E$ excludes missing data from the calculations. $\tilde{H}_E$ is calculated using the DeGiorgio & Rosenberg (2009) estimator, which accounts for the bias introduced by related individuals. $\hat{H}_E$ and $\hat{H}'_E$ are calculated using the Nei (1987) estimator, which does not take relatives into account. The legend applies to all three panels.

$\tilde{H}'_E$ across loci shows no decreasing trend over time (Fig. 1b), in contrast to the reported loss of variation over time seen for $\hat{H}_E$ by vonHoldt *et al.* (2008). The downward trend in $\hat{H}_E$ detected by vonHoldt *et al.* (2008) is instead likely to be caused by increasing average kinship in the sample after all founders had been introduced (Fig. 2). Additionally, as would be expected if inbreeding is rare, $\tilde{H}'_E$ and $\hat{H}'_O$ match more closely, both in value and in the lack of a temporal trend (Fig. 1b), than do the values of $\hat{H}_E$ and $\hat{H}_O$ (Fig. 1a) reported by vonHoldt *et al.* (2008). In fact, for each year of the study, considering paired lists of locus heterozygosities, we found $\tilde{H}'_E$ and $\hat{H}'_O$ not to be significantly different at the $P < 0.05$ level (Table 1). This similarity of $\tilde{H}'_E$ and $\hat{H}'_O$ and the absence of a downward temporal trend in these quantities are consistent with the low levels of inbreeding observed; these results are also compatible with the viewpoint of vonHoldt *et al.* (2008) that the population is thriving in terms of genetic diversity.

It is important to note that in our calculations of $\tilde{H}'_E$ and $\hat{H}'_O$, we treated the genotype data slightly differently from vonHoldt *et al.* (2008). In their computations of $\hat{H}_E$ and $\hat{H}_O$, missing data were treated as a separate allele. For a

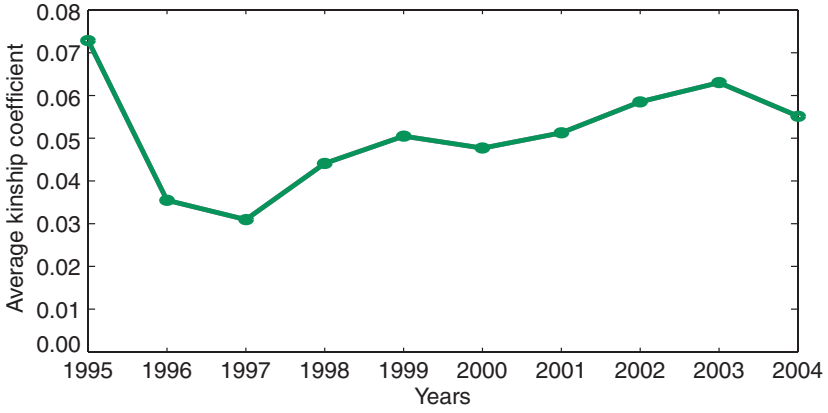

Fig. 2 Annual kinship coefficients averaged across all pairs of individuals genotyped for each locus, then averaged across all 26 loci. Individuals with missing data at a locus were excluded in $\bar{\Phi}$ computations at the locus for this plot and for the calculation of $\tilde{H}'_E$; they were not excluded in $\bar{\Phi}$ computations used in calculating $\tilde{H}_E$.

given individual at a given locus in the data of vonHoldt *et al.* (2008), data were always missing for both alleles or neither allele; therefore, treating missing data as an allele depresses $\hat{H}_O$ by increasing the proportion of 'homozygotes'. Comparing $\hat{H}'_O$ (Fig. 1b) to $\hat{H}_O$ (Fig. 1a), we can observe that the upward trend in $\hat{H}_O$ not observed for $\hat{H}'_O$ is partly explained by a difference in the treatment of missing data. As we will see below, however, this difference does not explain the difference in the trends seen for $\tilde{H}'_E$ (Fig. 1b) and $\hat{H}_E$ (Fig. 1a).

Treating missing data as a separate allele inflates $\hat{H}_E$, by adding another allele to the total number of distinct alleles in the calculation. Consequently, to ensure that the difference we observed between $\tilde{H}'_E$ (Fig. 1b) and the vonHoldt *et al.* (2008) estimates of $\hat{H}_E$ (Fig. 1a) was not the result of our differential handling of missing data, we compared annual values of $\tilde{H}_E$, obtained with the same approach to missing data as vonHoldt *et al.* (2008), to the previously reported values of $\hat{H}_E$ (Fig. 1c, top and centre lines). $\tilde{H}_E$, calculated with missing data counted as an allele, shows the same lack of temporal trend as $\tilde{H}'_E$, calculated with missing data excluded, and it differs from $\hat{H}_E$, calculated without accounting for relatives and including missing data as an allele. Additionally, the Nei (1987) estimator applied to samples with missing data excluded ($\hat{H}'_E$) shows a similar trend to the vonHoldt *et al.* (2008) values of $\hat{H}_E$, with missing data treated as a distinct allele (Fig. 1c, bottom and centre lines). We therefore conclude that the qualitative difference in expected heterozygosity we observe between the DeGiorgio & Rosenberg (2009) estimator and the vonHoldt *et al.* (2008) use of the Nei (1987) estimator is caused by differences in how the estimators treat relatedness, not in how missing data were handled.

In summary, using the unbiased DeGiorgio & Rosenberg (2009) estimator of expected heterozygosity with the Yellowstone grey wolves, we have determined that expected and observed heterozygosity are similar (Fig. 1b) and that indicators of genetic diversity do in fact correspond with behavioral observations of low inbreeding levels in the population. Our results also contrast with the previously published computations (vonHoldt *et al.* 2008) by finding no particular trend in expected heterozygosity over time. Additionally, whereas $\hat{H}'_E$ and $\hat{H}'_O$ differ significantly at the $P < 0.05$ level for seven of the 10 years of the study, the adjusted $\tilde{H}'_E$ matches $\hat{H}'_O$ more closely across all 10 years (Table 1). Thus, this example illustrates that the inherent bias in the standard Nei (1987) expected heterozygosity estimator caused by sampling of relatives can have a sizeable impact on estimated heterozygosity values and that the adjustment provided by the new DeGiorgio & Rosenberg (2009) estimator can alter the interpretation in cases in which relationships among individuals are largely known. As the Yellowstone wolves examined by vonHoldt *et al.* (2008) provide a prototypical genetic study of related individuals from a small natural population, our analysis suggests that the DeGiorgio & Rosenberg (2009) estimator will be informative in future analyses of the dynamics of gene diversity in the presence of close relatives.

Table 1 *P*-values for two-sided Wilcoxon signed-rank tests, comparing pairs of statistics across the 26 loci in the study

Year	$\hat{H}'_E$ vs. $\hat{H}'_O$	$\hat{H}'_E$ vs. $\tilde{H}'_E$	$\tilde{H}'_E$ vs. $\hat{H}'_O$
1995	0.0143	2.98×10^{-8}	0.5317
1996	0.3666	2.98×10^{-8}	0.5955
1997	0.3403	2.98×10^{-8}	0.7835
1998	0.2079	2.98×10^{-8}	0.4834
1999	0.0176	2.98×10^{-8}	0.9602
2000	0.0220	2.98×10^{-8}	0.9800
2001	0.0067	2.98×10^{-8}	0.8613
2002	0.0056	2.98×10^{-8}	0.9602
2003	0.0176	2.98×10^{-8}	0.9602
2004	0.0079	2.98×10^{-8}	0.7835

Acknowledgements

We thank M. Cronin and an anonymous reviewer for helpful comments and M. DeGiorgio for insight into $\tilde{H}_E$. Support for this work was provided by the Burroughs Wellcome Fund, the Alfred P. Sloan Foundation, and NIH grant R01 GM081441. Data are available upon request from B. M. vH. (bvonhold@ucla.edu).

References

Atkinson B, Therneau T (2008) kinship: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. R package version 1.1.0-22.

Bourgain C, Abney M, Schneider D, Ober C, McPeek MS (2004) Testing for Hardy–Weinberg equilibrium in samples with related individuals. *Genetics*, **168**, 2349–2361.

Broman KW (2001) Estimation of allele frequencies with data on sibships. *Genetic Epidemiology*, **20**, 307–315.

DeGiorgio M, Rosenberg NA (2009) An unbiased estimator of gene diversity in samples containing related individuals. *Molecular Biology and Evolution*, **26**, 501–512.

Lange K (2002) *Mathematical and Statistical Methods for Genetic Analysis*, 2nd edn. Springer-Verlag, New York.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

vonHoldt BM, Stahler DR, Smith DW, Earl DA, Pollinger JP, Wayne RK (2008) The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Molecular Ecology*, **17**, 252–274.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Ivana Jankovic is interested in the analysis of population-genetic data and has worked on method development and applications. Bridgett vonHoldt studies evolutionary genomics in domestic and wild canids, and epigenetic modifications of transposable elements in domestic plant species. Research in Noah Rosenberg's laboratory focuses primarily on mathematical theory and statistical methods in population genetics.

---

doi: 10.1111/j.1365-294X.2010.04746.x