

[MBC out of reprints]

Journal of Heredity

JOURNAL HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS

QUICK SEARCH: [advan

Author:	Keyword(s):
Go	
Year:	Vol: Page:

The Journal of Heredity 2002:93(4)
 © 2002 The American Genetic Association 93:260-269

Comparisons of Likelihood and Machine Learning Methods of Individual Classification

B. Guinand, A. Topchy, K. S. Page, M. K. Burnham-Curtis, W. F. Punch, and K. T. Scribner

From the Departments of Fisheries and Wildlife (Guinand, Page, and Scribner) and Computer Science and Engineering (Topchy and Punch), Michigan State University, East Lansing, MI 48824; and the USGS Great Lakes Science Center, 1451 Green Rd., Ann Arbor, MI 48105 (Burnham-Curtis). Bruno Guinand is currently at UMR CNRS 5000 Génome, Populations, Interactions, Station Méditerranéenne de l'Environnement Littoral, 1, Quai de la Daurade, F34200 Sète, France. Kevin S. Page is currently at the Minnesota Department of Natural Resources, Division of Fisheries, 1601 Minnesota Dr., Brainerd, MN 56401. Mary K. Burnham-Curtis is currently at the U.S. Fish and Wildlife Service, National Fish and Wildlife Forensics Laboratory, 1490 East Main St., Ashland, OR 97520.

Address correspondence to Kim T. Scribner at the address above, or e-mail: scribne3@pilot.msu.edu.

Classification methods used in machine learning (e.g., artificial neural networks, decision trees, and k -nearest neighbor clustering) are rarely used with population genetic data. We compare different nonparametric machine learning techniques with parametric likelihood estimations commonly employed in population genetics for purposes of assigning individuals to their population of origin ("assignment tests"). Classifier accuracy was compared across simulated data sets representing different levels of population differentiation (low and high F_{ST}), number of loci surveyed (5 and 10), and allelic diversity (average of three or eight alleles per locus). Empirical data for the lake trout (*Salvelinus namaycush*) exhibiting levels of population differentiation comparable to those used in simulations were examined to further evaluate and compare classification methods. Classification error rates associated with artificial neural networks and likelihood estimators were lower for simulated data sets compared to k -nearest neighbor and decision tree classifiers over the entire range of parameters considered. Artificial neural networks only marginally outperformed the likelihood method for simulated data (0–2.8% lower error rates). The relative performance of each machine learning classifier improved relative likelihood estimators for empirical data sets, suggesting an ability to "learn" and utilize properties of empirical genotypic arrays intrinsic to each population. Likelihood-based estimation methods provide a more accessible option for reliable assignment of individuals to the population of origin due to the intricacies in development and evaluation of artificial neural networks.

This Article

- ▶ [Full Text](#)
- ▶ [Full Text \(PDF\)](#)
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in ISI Web of Science](#)
- ▶ [Similar articles in PubMed](#)
- ▶ [Alert me to new issues of the journal](#)
- ▶ [Download to citation manager](#)
- ▶ [Search for citing articles in: ISI Web of Science \(3\)](#)
- ▶ [Request Permissions](#)

PubMed

- ▶ [PubMed Citation](#)
- ▶ [Articles by Guinand, B.](#)
- ▶ [Articles by Scribner, K. T.](#)