Tutorial section

WinPop 2.5: Software for representing population genetics phenomena

Abstract
The curriculum for genetics courses is shifting from a classical to a more molecular genetics focus, increasing the importance of subjects such as population genetics. Population genetics is a computational and statistical field that requires a good understanding of the nature of stochastic events. It is a difficult field for biology students with a limited mathematical background and there is a need for visualisation tools to facilitate understanding by the use of practical examples. WinPop provides students and researchers with a visual tool to allow the simulation and representation of population genetics phenomena. WinPop is a user-friendly software meant for use in population genetics courses and basic research. WinPop 2.5 contains six different modules that represent and simulate population genetics models. Genotype and allele frequencies are calculated under the different models: panmixia, genetic drift, assortative matings, selection, gene flow and mutation. The program’s interface presents information in Cartesian graphics and isosceles triangular coordinate systems, allowing the user to save graphical and textual data output from the simulations. WinPop is developed in Visual Basic 6.0 and uses Windows 95 and higher. WinPop 2.5 can be downloaded from http://www.genedrift.org/winpop.php

INTRODUCTION
Population genetics provides the theoretical background behind genetics and molecular biology and is thus fundamental to a curriculum in the Biological Sciences. Additionally, according to Kliman,¹ (1) a proper understanding of microevolutionary processes requires understanding of population genetics beyond Hardy–Weinberg (HW) equilibrium; (2) population genetics offers opportunities to generate and test hypotheses using quantitative methods; and (3) population genetics is one of the few subjects that introduce students to stochastic problems. An issue in population genetics teaching is to present data in a visually appealing manner to simplify the understanding of different phenomena, providing students and professors with a graphical output of the formulae and equations. With the creation of faster and more powerful computer systems, the development of useful tools to study and teach population genetics has become a straightforward task. Nuin and Otto² released WinPop v. 1.0, software developed for Windows 3.x systems using a 16 bit version of Visual Basic 3.0. WinPop 2.5 presented in this work is an extension of this previous version, including new modules, faster simulations and several new graphical features.

The WinPop 2.5 interface is designed as a Multiple Document Interface (MDI) with different children windows for each phenomenon simulated. Classical Cartesian coordinate systems and an isosceles triangular coordinate system³ are used to represent the different population genetics models, making WinPop a

Keywords: population genetics, simulation, panmixia, genetic drift, Markov chains
valuable visual tool to introductory and advanced population genetics courses.

**DISCUSSION**

WinPop 2.5 generates representations or simulations of six different population genetics phenomena. It allows calculations of genotype and allele frequencies for models of panmixia, genetic drift, assortative mating, selection, gene flow and mutation. Each module is presented in its own MDI child window with text boxes for value input, a white canvas that handles the graphical output (where needed) and command buttons to control usability. The access to each module’s window is done through a menu on the MDI parent frame. All graphs generated can be saved as a regular Windows Bitmap file (*.BMP), while the resulting (and in some cases partial) numerical values from the calculations can be stored in tab-delimited ASCII files.

The panmixia module consists of representations of a single diallelic autosomal locus frequencies inside the isosceles triangular coordinates, a diallelic X-linked loci in a Cartesian graph, simple demonstrations of genotypic and allele frequencies and \( \chi^2 \) calculations. In order to introduce the triangular coordinates system, there is also an interactive graph, where any mouse click inside the isosceles triangle displays all the five frequencies represented by each point. This is one of the advantages of using the triangular coordinates to represent frequencies, as any plotted point contains all the information about the population. Also in this module, there is the possibility of representing allelic frequencies of a diallelic X-linked locus,\(^6\) starting with the input of allele frequencies among males and females in one population. WinPop displays the frequency curves in Cartesian coordinates and also gives visual information about the numerical values of the genotypic frequencies.

In the Assortative Matings module the user is able to select a complete set of models for this type of non-random mating. The program also takes advantage of the triangular coordinates system to plot allelic frequencies points along a user set number of generations. Also, in this case a diallelic locus is considered, and it can be studied under several models: positive assortative matings (with and without dominance), negative assortative matings (disassortative matings),\(^7\) sterile and forbidden assortative matings.\(^8,9\) The first two are a predefined set of crossings that are occurring (or not) in the population, while in the latter the user selects from a crossing matrix the matings that would be sterile and forbidden. The allelic frequencies calculated can be plotted in Cartesian graphs and there is an option to generate equilibrium points (stable and unstable) for each set of matings.

In the Natural Selection module, it is possible to represent a general model of selection and also a frequency-dependent selection model, both using the triangular coordinates.\(^7\) When studying the general model of selection, a \( \Delta q \) graph can be obtained, displaying the allelic frequency equilibrium points. All calculated values can be output in Cartesian graphs. Selection can be studied in combination with genetic drift.

The genetic drift module is a simulation of allelic frequencies along generations at a single diallelic locus.\(^10\) The user has full control of the input values: population size, number of generations, number of populations and initial frequencies. In addition fitness coefficients for each one of the genotypes and mutation rates (uni- or bidirectional) can also be set, making this simulation extremely useful to understand frequency variation in populations under different influences. All simulated values, final and partial, can be saved in an ASCII format file.

Another aspect of genetic drift covered in WinPop, are the Markov chains that represent expected values of this phenomena (such as the Wright–Fisher model).\(^6\) A population size up to 70 individuals can be displayed in this module. This limitation is due to screen
size constraints and Visual Basic graphical object dimensions. After entering the population size, all possible allele frequency values are shown in a dropdown box and the user can select any of these values as the starting frequency. The partial and final outputs are displayed as a numerical vector of probabilities and as a bar graph representing these values. Markov chains are notorious for being particularly challenging among students in introductory population genetics courses, and this user-friendly module is an excellent visual aid.

Mutation has a separate module, where uni- or bidirectional mutation pressures on allelic frequencies can be plotted in a Cartesian graph. Different curves can be displayed in the same graph, making this simple module useful in the comparison of the effects of different mutation rates on the allelic frequencies with time.

WinPop permits the possibility of studying the linkage disequilibrium between two diallelic loci. In this module the coefficient of disequilibrium ($D$) can be calculated from the haplotype absolute frequencies input values. The program calculates $\chi^2$ values for the sample and plots a bar graph with a comparison of the observed and expected values. The magnitude of $D$ based on different recombination rates can be plotted in a Cartesian graph, displaying the decay of $D$ caused by recombination among the haplotypes.

Lastly, WinPop represents in a Cartesian graph, allelic frequencies for a gene in two populations under the effects of gene flow. The migration rates input by the user can be uni- or bidirectional. In this module the allele frequencies with different migration rates can be compared in the same graph. Every WinPop module has a complete ‘how-to’ help file with guides on how to enter input values, start representations/simulations and to generate output as a file. Included in this help file are sample exercises that can be solved using the program’s representations. Every MDI child form in WinPop has a ‘What’s this?’ help button. When clicked, descriptions are given for the objects on the screen (tooltips). These ‘tips’ guide the user to what values should be entered and where, and give a description of what is being output by the program.

CONCLUSIONS
WinPop 2.5 is intended for population genetics courses and very basic population genetics research, providing students and researchers with a powerful visual tool that allows representations/simulations of common population genetics phenomena. The available hypertext help file is complete with program tutorial and some suggested exercises. Also there are visual aids such as the isosceles triangular coordinates explanation and tooltips for the objects on screen. The software is available free of charge, can be installed in Windows OS System version 95 and above, and is compatible with any Intel-like computer system.

Acknowledgments
I would like to thank Brian Golding for comments on the software and manuscript, Elisabeth Tillier for comments and suggestions on the manuscript and Paulo A. Otto for the inspiration to create the software and the students from the 2004 Population Genetics course 3J03 at McMaster that helped to improve the program.

Paulo A. S. Nuin
Division of Genomics and Proteomics in Cancer,
Ontario Cancer Institute,
University Health Network,
620 University Ave,
Toronto, ON, Canada M5G 2M9

References


