PyPedal, an open source software package for pedigree analysis

John B. Cole, Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD, USA

The open source software package PyPedal (http://pypedal.sourceforge.net/) was first released in 2002, and provided users with a set of simple tools for manipulating pedigrees. Its flexibility has been demonstrated by its used in a number of settings for large and small populations. After substantial revisions and feature additions, an alpha version of PyPedal 2 became available in 2006. The production version of PyPedal 2.0.0 was released in 2010, and featured a completely rewritten object model and many tools for pedigree manipulation and analysis. Many measures of genetic variation can be calculated from pedigrees, including effective population sizes; effective founder and ancestor numbers; and coefficients of regular, ancestral, and partial inbreeding. The software has been used on pedigrees of up to 600,000 animals in several species, including dogs, dairy cattle, and beef cattle. Data can be loaded from, and saved to, plain-text files, GEDCOM 1.5 and GENES 1.20 binary files; and MySQL, Postgres, SQLite and databases. Version 2.1.0, which is currently undergoing beta testing, adds several new tools for manipulating pedigrees by treating them as sets. Pedigrees can be merged by taking the union of two or more pedigrees, for example. The intersection of a two or more pedigrees is the set of animals common to all of them, and union and intersection operations can be used together to perform operations analogous to subtraction. Fast algorithms for computing coefficients of inbreeding also have been added. Unknown parent groups can be constructed based on a description provided by the user. Missing values are now handled more consistently, and missing values within individual records can be accommodated. Support for a data frame class provides a much more efficient framework for calculating summary statistics, such as demographic measurements. PyPedal is provided under an open source license, and may be freely modified, distributed, and used.