PROGRAM NOTE

GMCONVERT: file conversion for GENEMAPPER output files

BRANT C. FAIRCLOTH
D. B. Warnell School of Forestry and Natural Resources, The University of Georgia, Athens, Georgia, 30602, USA

Abstract

GMCONVERT **is a platform-independent program provided in GUI (for Apple OS X and Windows XP) and command-line versions (for other platforms).** GMCONVERT **allows rapid reformatting of microsatellite data from output files produced by Applied Biosystems** GENEMAPPER **software (version 3.x). The program will re-array data into three formats commonly used in downstream analysis:** GENEPOP, CERVUS, **and** GERUD. GMCONVERT **will greatly increase the speed of data preparation prior to analysis and aid in reducing transpositional errors associated with manual re-arraying and reformatting steps.** GMCONVERT **is available from http://gallus.forestry.uga.edu/software/.**

Keywords: data conversion, GENEMAPPER, microsatellites, population genetics, PYTHON

Received 8 February 2006; revision received 16 March 2006; accepted 31 March 2006

Analysis of microsatellite data frequently requires conversion from one format to another for use by various software packages. Manual data conversion and re-arraying is time-consuming and potentially error-prone, particularly when sample sizes or numbers of loci are large. Several programs can convert data between various program formats (Marshall *et al*. 1998; Dieringer & Schlötterer 2003; Glaubitz 2004). However, most current programs assume the data are already in some 'common' (e.g. conversion program dependent or GENEPOP) format. Few programs exist to convert data from the formats output by or exported from fragment analysis software packages (e.g. GENEMAPPER) distributed with common genotyping/sequencing instruments (Applied Biosystems 3100, 3130, and 3730).

I wrote GMCONVERT to enable rapid and accurate conversion and re-arraying of data exported from GENEMAPPER prior to downstream analysis. GMCONVERT may be downloaded as a disk image (for OS X), a zip archive (for Windows), or a gzipped, tar archive at the following hyperlink: http://gallus.forestry.uga.edu/software/.

GMCONVERT is written in PYTHON (Python Software Foundation 2006) and is released under the GNU General Public License (GPL). PYTHON is an open-source, interpreted, object-orientated programming language distributed for multiple platforms (Windows, Linux, and Unix-like) and comparable to Perl, Java, or Ruby. PYTHON is available as an easily installed, binary package for most operating systems.

Correspondence: Brant C. Faircloth, Fax: 706-542-8356; E-mail: brant@uga.edu

PYTHON may also be compiled from source code. Graphical user interface (GUI) versions of GMCONVERT make use of the wxPython package (available from http://wxpython.org/).

I provide binary, executable GUI versions (for Microsoft Windows and Apple OS X; Fig. 1) and a command-line version of GMCONVERT, allowing batch processing or use on other platforms. The executable versions of GMCONVERT do not require that the PYTHON language be installed on the client machine prior to running the program. The command line version requires that PYTHON be installed before the program can be run.

GMCONVERT takes as input text files output by the GENEMAPPER 'export' command (select 'Genotypes' tab; then File > Export Table). It is assumed that exported data have been binned and reviewed for errors and are ready for downstream analysis. Exported data must retain at least the following columns for proper conversion: Sample File, Sample Name, Marker, Allele 1, Allele 2, and GQ (genotype quality). Marker name is case-sensitive during the conversion process, and users should ensure names of identical loci within different GENEMAPPER panels are consistent. GMCONVERT will filter out those samples with GQ scores < 0.75 at each locus. The program will then re-array data into one of three, user-selected formats: GENEPOP (Raymond & Rousset 1995), CERVUS (Marshall *et al*. 1998) and GERUD (Jones 2005). Samples with no allelic value at a locus will be marked with the null character correct for the output file format chosen. These values will be converted to 'Nulls Not Allowed!' for the GERUD format in which null values are not allowed. Any alleles that remain
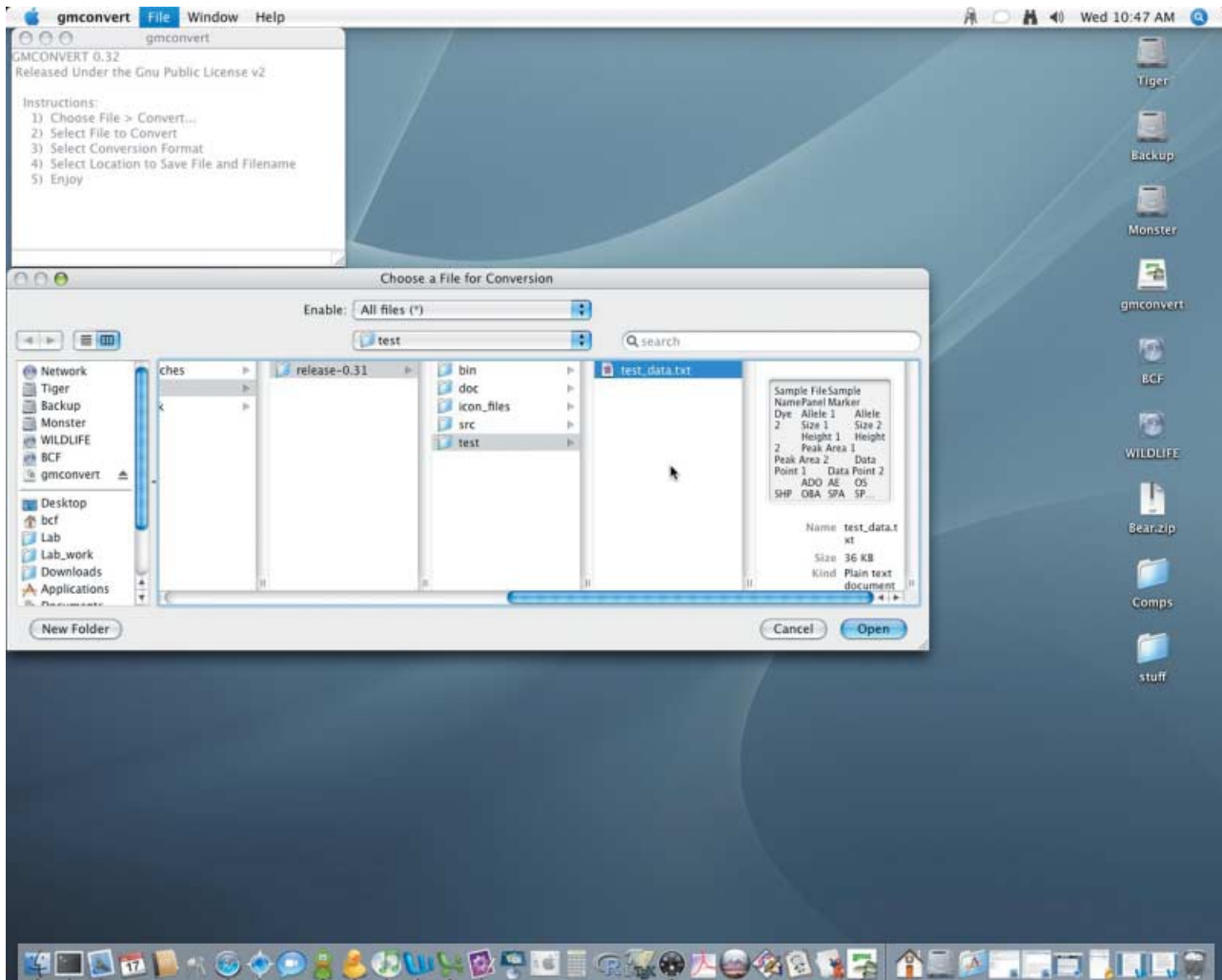
Fig. 1 GMCONVERT graphical user interface on the OS × platform.

out of bin prior to conversion will be marked with '!!!Bin!!!' in the output file.

GMCONVERT has been tested with GENEMAPPER files of several hundred samples genotyped at > 10 loci. Functionality and speed of the program are theoretically limited by the processor type and speed, the amount and speed of RAM installed on the system, the operating system, the version of PYTHON installed, and the size of the input file. The GUI versions of GMCONVERT have been tested successfully on Apple OS X (10.4.5 and 10.3.9) and Windows XP (Service Pack 2). The command line version has been tested on OS X (10.4.5), Windows XP Service Pack 2, FreeBSD 5.4, and Linux (Gentoo, kernel 2.6.14).

The documentation distributed with GMCONVERT provides GPL license and warranty terms in addition to a README file containing information on installing and using GMCONVERT and a sample data file. GMCONVERT will be maintained and upgraded for the foreseeable future,

and additional output formats may be added upon request or via modification of the source code.

Acknowledgements

References

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Glaubitz JC (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, **4**, 309–310.

Jones AG (2005) GERUD 2.0: a computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. *Molecular Ecology Notes*, **5**, 708–711.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Python Software Foundation (2006) PYTHON version 2.4.2. Available from: http://www.python.org.