

Ian A. Dickie · Peter G. Avis · David J. McLaughlin ·
Peter B. Reich

Good-Enough RFLP Matcher (GERM) program

Received: 7 August 2002 / Accepted: 7 January 2003 / Published online: 11 February 2003
© Springer-Verlag 2003

Abstract A spreadsheet-based program (Good-Enough RFLP Matcher or GERM) is presented that matches unknown restriction fragment length polymorphism (RFLP) patterns of ectomycorrhizal fungi to a database of known ectomycorrhizal fungi. The program uses three simple methods to determine whether a sample matches a known: (1) Forward Matching: whether every band in the unknown is present in a known sample within a given error range; (2) Backward Matching: whether every band in the known sample is present in the unknown within a given error range; (3) Sum of Bands: whether the sum of all bands in the known and unknown are similar within a given error range. The program is available through the web page of this journal.

Keywords Ectomycorrhiza · Fungi · Identification

Introduction

Restriction fragment length polymorphism (RFLP) analysis has been widely adopted by mycorrhizal researchers as a powerful tool for the identification of ectomycorrhizal root tips (Horton and Bruns 2001). To utilize RFLP analysis, RFLP patterns from unknown ectomycorrhizal root tips must be matched with patterns from known ectomycorrhizal fungal tissue. Known ectomycorrhizal

Electronic Supplementary Material Supplementary material is available for this article if you access the article at <http://dx.doi.org/10.1007/s00572-003-0225-x>. A link in the frame on the left on that page takes you directly to the supplementary material.

I. A. Dickie (✉) · P. B. Reich
Department of Forest Resources,
University of Minnesota,
115 Green Hall, 1530 N Cleveland Ave., St. Paul, MN 55108, USA
e-mail: dicki017@umn.edu
Tel.: +1-612-6244916, Fax: +1-612-6255212

P. G. Avis · D. J. McLaughlin
Department of Plant Biology,
University of Minnesota,
1445 Gortner Ave., St. Paul, MN 55108, USA

samples are often obtained from sporocarps or from root tip samples that have been sequenced. In relatively low-diversity ecosystems, the matching of RFLP patterns can be accomplished by visual examination. However, in ecosystems with high ectomycorrhizal diversity, the use of computerized RFLP pattern matching programs becomes necessary. A number of commercial programs are available that will accomplish this, but the cost of these programs may be excessive for many researchers.

Recently, we have developed a spreadsheet-based program to match unknown RFLP patterns to a database of known samples. This program, named “Good-Enough RFLP Matcher” or “GERM”, uses three complementary methods to determine whether a sample has a match in the database (see Methods). GERM is based in Microsoft Excel (Microsoft, Redmond, Wash., USA) and uses both normal spreadsheet formulas and macros programmed in Visual Basic for Applications (VBA). GERM is being made available through the web site of this journal in both a full and limited version (the limited version may function better on computers with limited memory). The full version (GERM 1.01) can be used with up to four RFLP enzymes, 400 known samples in the database and 10 distinct bands for each enzyme. The limited version (GERM_LE 1.01) can be used with up to four RFLP enzymes, 200 known samples in the database, and seven distinct bands for each enzyme. In either version, multiple databases can be maintained to increase the number of samples that can be matched. The program will work on any computer that supports Excel 2000 (PC) or Excel 98 (Macintosh). Macintosh users with Excel 98 will need to increase memory allocation to Excel to use the program (to 19,000 K for the full version, or 10,000 K for the LE version). Full documentation for the program is included within the Excel worksheet.

Methods

The program uses three methods to determine whether an unknown sample matches an entry in the known database. The Forward Matching method checks whether every band in the unknown

sample has a matching band in the known sample within a user-defined error limit. This is calculated using the array formula: $\{=MIN(ABS([unknown\ band\ length]-[array\ of\ all\ band\ lengths\ for\ a\ given\ known]))\}$ repeated for every unknown band length (1 to 10 in the full version), for every known (1 to 400 in the full version), and for each enzyme (1 to 4 in both versions). Formula names follow Excel: MIN = minimum, ABS = absolute value. The {} brackets are Excel nomenclature for an array formula, permitting an array of band lengths to be subtracted from a single value. Text in [] brackets indicates references to data cells. The forward error is calculated as the maximum error for any band within the unknown.

The Backward Matching method checks whether every band in each known sample has a matching band in the unknown sample. Similarly to the Forward method, this is calculated with the array formula: $\{=MIN(ABS([known\ band\ length] - [array\ of\ all\ band\ lengths\ in\ the\ unknown]))\}$ for all known band lengths for each known and for all enzymes, with the maximum error for any band in the known being reported.

The Sum method compares the sum of all of the bands in the known with the sum of all of the bands in the unknown and determines whether the difference is less than a user-defined error. This is the formula: $=ABS(SUM[array\ of\ known\ band\ lengths] - SUM[array\ of\ unknown\ band\ lengths])$. The Sum method helps to avoid matching a single band to multiple bands, which can occur if the first two methods are used alone.

Data entered into GERM can include faint or uncertain bands. In our experience, we have sometimes obtained RFLP patterns in which one or more bands are faint or where a single band is believed to be two superimposed bands that are not resolvable. In either of these cases, the faint or the second of a suspected double band pair can be entered in the program, preceded by a “-” sign to indicate that the band may or may not be present. In addition, the program automatically treats low-bp bands (within one error margin of a lower threshold set by the user) as suspect bands; these low-bp bands are likely to be absent in some replicate samples and are, therefore, unreliable to distinguish species. Suspect bands are included in matches where they permit a match, but are not used to exclude possible matches. Where either the unknown sample or the known sample contains a faint or low-bp band, the Sum method is disabled.

Error thresholds for each of the three methods are calculated as a standard number of base pairs. Examination of replicate RFLP samples run on different gels suggested that this was an appropriate measure of error, rather than using percentage of band length (Dickie and McLaughlin, unpublished data).

The program ranks possible matches by one of four methods. The default ranking method is Joint, which is the sum of errors (in bp) from the forward and backward matching methods. Ranking can also be based on any one of the three matching methods independently (Forward, Backward, or Sum). The top 50 matches from the database are reported, although the degree of match for every sample in the known database can also be examined. In addition to a text report of the highest ranked 50 matches from the database, the 10 highest ranked matches are graphically shown.

GERM will base matches on as many or few RFLP enzymes for which there are data for both the unknown and any given known. The number of enzymes on which a match is based is reported.

Discussion

The most appropriate error level for acceptance of a match will need to be determined by each laboratory based on the reproducibility of their results. Setting the error too high will result in spurious matches, while too low a setting will result in missing possible matches. Our

preference has been to set the error relatively high (25 bp error for Forward and Backward methods, 100 bp error for Sum method) and then to compare suggested matches visually. The lower threshold for measurement will depend on the ladder used. We normally set this value to the lowest band in the ladder used (100 bp using our standard ladder). Following Gardes and Bruns (1996), we suggest that matches only be considered if at least two enzymes have been used.

Database management becomes very important when dealing with large numbers of samples. GERM can contain an unlimited (within the memory limits of the computer) number of databases of known species, with up to 400 samples in each database (200 in the LE version). We suggest that separate databases be maintained for sporocarps and for unknown samples. When new sporocarp or root tip samples are added to the database, they can be run against the database of unknown samples first to identify any new matches. This will obviate the need to re-run the program for every unknown sample every time a new known sporocarp is added.

This program is intended as a tool to suggest matches. Visual inspection of proposed matches remains vital, particularly if one or more of the matching methods is disabled or if acceptable error values are set high. We view a match as more reliable if the errors of measurement are similar for all bands, and less reliable if some bands are mismatched by being too long, and others too short.

The program contains a sample database, which comprises data collected by D. McLaughlin and associates from an oak savanna site near St. Paul, Minn., USA (McLaughlin et al., unpublished data). This data is supplied for illustrative purposes. Because of high levels of intraspecific variation in the ITS region (Horton 2002), it is critical that independent databases be developed for each site of interest.

Acknowledgements T. Horton assisted with testing and improving this program. Major support was provided by an NSF LTER grant (NSF/DEB 0080382). This research was also supported in part by the University of Minnesota Agricultural Experiment Station.

References

- Gardes M, Bruns TD (1996) ITS-RFLP matching for the identification of fungi. In Clapp JP (ed) *Methods in molecular biology, species diagnostics protocols: PCR and other nucleic acid methods*. Humana, Totowa, N.J., 50:177–186
- Horton TR (2002) Molecular approaches to ectomycorrhizal diversity studies: variation in ITS at a local scale. *Plant Soil* 244:29–39
- Horton TR, Bruns TD (2001) The molecular revolution in ectomycorrhizal ecology: peeking into the black box. *Mol Ecol* 10:1855–1871