MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) **13**, 726–733 doi: 10.1111/1755-0998.12094

Analysing diversity and community structures using PCR-RFLP: a new software application

FABIENNE FLESSA,* ALEXANDRA KEHL† and MATTHIAS KOHL‡
*Section Mycology, University of Bayreuth, 95440 Bayreuth, Germany, †Botanical Garden, University of Tuebingen, 72076 Tuebingen, Germany, ‡Department of Medical and Life Sciences, Furtwangen University, 78054 Villingen-Schwenningen, Germany

Abstract

Restriction fragment length polymorphism tools is an R application which supports a complete workflow of polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), dealing with the problems which accompany analysis when PCR-RFLP is used in diversity studies. Large numbers of different RFLP samples obtained from multiple electrophoresis runs might lead to limitations or misidentifications due to the need for band matching in most existing software applications. Due to the common problem of variation in the density of bands (i.e. distances between bands or visual intensity) in the electropherograms, it is desirable to have options for handling samples with uncertain or faint bands. As a further step in the workflow, scientists often use DNA sequencing to identify individual genotypes, so that the use of specific software to combine these tasks might be helpful. With this background, we here present an application that supports a complete workflow, starting with the analysis of single species samples by PCR-RFLP, to PCR-RFLP genotype identification based on a reference data set and DNA sequencing followed by similarity analysis. RFLPtools is a freely available, platform-independent application which provides analysis functions for DNA fragment molecular weights (e.g. by RFLP analysis), including similarity calculations without the need for band matching. As it is written for the statistical software R, other statistical analyses might also be easily applied.

Keywords: diversity studies, R package, RFLP analysis, sequence analysis

Received 30 August 2011; revision received 19 January 2013; accepted 31 January 2013

Introduction

In many biological disciplines, fingerprinting techniques such as restriction fragment length polymorphism (RFLP) have become useful and commonly applied laboratory tools in biodiversity research to recognize and analyse genotypic diversity. The RFLP technique, especially, is frequently used in population studies targeting various organisms such as plants (Tsarouhas *et al.* 2002) or symbiotic (Sykorova *et al.* 2007), pathogenic, or soil-inhabiting microbes (Watrud *et al.* 2006; Duran *et al.* 2009). Polymerase Chain Reaction (PCR)-RFLP is based on nucleotide differences that affect the binding site for restriction enzymes in certain DNA arrays, which after enzymatic digestion and amplification of the fragments lead to different fragment patterns following visualization by electrophoresis. Other related fingerprinting techniques are Amplified Fragment Length Polymorphism (AFLP) (Vos *et al.* 1995), Random Amplification of Polymorphic DNA (RAPD), STRs (Single Tandem Repeats), microsatellites (Goldstein & Schlötterer 1999) and Terminal RFLP (T-RFLP) (Liu *et al.* 1997).

By comparing the resulting fragment patterns, the RFLP technique is capable of detecting and discriminating genotypes after a cloning step to gain single species samples from environmental samples or populations. To further characterize genotypes (or organisms) in such studies, PCR products of single RFLP samples can be DNA-sequenced and the resulting sequences further compared by nucleotide sequence alignment searches performed by specific algorithms, for example by the standalone Basic Local Alignment Search Tool (BLAST) (Zhang *et al.* 2000) or local BLAST v7.0.9 (Hall 1999), with published sequence data.

However, sample treatment and the analysis of fragment patterns are susceptible to technical difficulties if electrophoresis gels are used, which might affect and reduce the reproducibility and reliability of results. The processing of fragment patterns might be hampered, for example, by more or less serious lateral deformations of runs during electrophoresis (so-called 'smiley effects') or

Correspondence: Matthias Kohl, Fax: +49-7720-307-4207;
E-mail: Matthias.Kohl@stamats.de

by the presence of large amounts of highly diverse and sometimes unresolved or superimposed bands, which contribute to analysis and interpretation bias. Additionally, the comparability of large numbers of electrophoretic runs is sometimes limited, especially in biodiversity studies, where large amounts of samples are usually generated on numerous electrophoresis gels. In the analysis and comparison of the resulting fragment patterns, band matching is a precondition for subsequent analyses in most software applications and is often problematic due to the mentioned constraints. Furthermore, proprietary applications that provide adequate analysis functions are mostly rather expensive.

Comparisons of fragment sizes or peak heights are already implemented in free bioinformatic software applications such as GERM (Dickie *et al.* 2003), which is based on Visual Basic application macros, or FragMatch, a Java application (Saari *et al.* 2007). Table 1 summarizes the possibilities and limitations of these two applications compared with RFLPtools. Free software tools for other genome fingerprinting techniques also exist (e.g. Genographer, RawGeno, OptiFLP). However, tools for RFLP analysis disregard some of the specific problems which occur during the analysis of electrophoresis runs, such as biased electrophoresis fragment patterns and deformations during gel runs such as smiley effects. The problems often propagate as the number of samples analysed, and the diversity of band patterns increase.

Therefore, a free and platform-independent method for reading fragment pattern data derived from molecular fingerprinting analysis such as PCR-RFLP, without the constraints of matching bands in parallel runs, is required. Using the methods of matching bands, it is essential that there is no deformation in the electrophoresis gel. As soon as nonhomogeneous flow rates exist, it is difficult to define the exact distance for each fragment in the run. It is most probably that identical samples are interpreted as different RFLP types if the method of matching bands is applied to the run. Therefore, in RFLPtools, instead of band matching, distances are used for the analysis of the RFLP types.

Restriction fragment length polymorphism tools was mainly designed for data gathered from images of electrophoresis gels. Due to the common problem that band density and resolution might vary between individual gels, options to handle samples with doubtful or faint bands are desirable. To improve data quality, it is also necessary to have options for checking the reliability of the data.

Particularly for diversity studies with large and heterogeneous data sets, such analysis software should include a function for comparing fragment patterns with a reference data set comprising known fragment patterns of previously identified genotypes or organisms. Such an option should allow the fast matching of newly recognized samples with already characterized RFLP samples. A tool for the visual comparison of fragment patterns with reference data sets would help facilitate the rapid identification of unknown samples. For further processing of the data obtained, a suitable, freely available analysis application with a high degree of interoperability and flexibility and the option to read and write standard data exchange files for communicating with applications that provide additional statistical analysis methods is still needed.

Table 1 Comparison of GERM, FragMatch, and RFLPtools

	GERM	FragMatch	RFLPtools
System requirements	Microsoft Excel	Java	Independent
Based on	Spreadsheet formulas and macros, programmed in Visual Basic	Java	R
Quality check for input data	No	No	Yes
Different methods for analysis of RFLP fragment patterns	No	No	Yes
Method for analysis of RFLP fragment patterns	Band matching	Band matching	Distances
Methods for handling biased data	No	No	Yes
Methods for handling doubtful bands at a lower length threshold	Yes	No	Yes
Graphical visualization	Yes	Yes	Yes
Ability to compare new data with reference database	Yes	Yes	Yes
Designed for multiple data sets from a single sample	Yes	Yes	Yes
Function for comparing sequence similarities	No	No	Yes

Methods and implementation

Restriction fragment length polymorphism tools was developed to support the complete workflow, from the

analysis of biotic diversity after a cloning step, to obtaining single species samples from environmental samples by PCR-RFLP and to the identification of RFLP fragment patterns via sequencing. It includes several functions to import fragment pattern data, estimate their similarity and compare the resulting RFLP types with a reference data set for fast identification, as well as the possibility of a similarity analysis of DNA sequences.

In this section, we explain the functions and their use during a biodiversity study.

Restriction fragment length polymorphism tools, like many other R packages (R Development Core Team 2012), is command-line driven. It is installed by the command install.packages('RFLPtools', dependencies = TRUE). Functions are directly invoked by the user, in parts specified by arguments and options. Every session using RFLPtools in R starts with the command 'library (RFLPtools)', which loads all RFLPtools functions into the R environment. Information about the package (i.e. the DESCRIPTION file), as well as the list of included functions, can be displayed with the command 'library (help = RFLPtools)' (see also Table 2). There is a detailed help file (vignette), showing typical workflows which can be opened inside R by vignette('RFLPtools').

Before starting the analysis of fragment patterns via RFLPtools, fragments have to be detected using any graphical fragment pattern analysis software package, such as GeneProfiler (Scanalytics Inc.) or the free gel analysis macro MolWt (http://www.phase-hl.com/imagej.htm), which are able to generate simple text report files, including data on the molecular weight of detected bands, sample names and band numbers.

Table 2 Functions and their description

Name	Description
RFLPcombine	Combine multiple data sets from a single sample (i.e. separate digests with two or more enzymes)
RFLPdist	Compute distances for RFLP data
RFLPdist2	Compute distances for RFLP data where some bands may be missing
diffDist	Distance matrix computation based on successive differences
linCombDist	Linear combination of distances
RFLPdist2ref	Compute distance between RFLP data and RFLP reference data
RFLPlod	Remove bands below lower length threshold
RFLPplot	Plot RFLP data
RFLPqc	Quality control for RFLP data
RFLPrefplot	Function for a visual comparison of RFLP samples with reference samples
nrBands	Function to compute number of bands
read.blast	Read BLAST data
read.rflp	Read RFLP data
sim2dist	Convert similarity matrix to dist object
simMatrix	Similarity matrix for BLAST data
write.hclust	Cut a hierarchical cluster tree and write cluster identifiers to a text file

In contrast to existing applications, RFLPtools provides an optional quality check of the RFLP band patterns by comparing the sum of the molecular weights of all bands per sample with the given molecular weight of the complete PCR product within a certain range of tolerance (function RFLPqc). This option allows the identification of biased data and provides decision support for either excluding the respective samples from further analysis or for rechecking the doubtful band patterns on the original gel electrophoresis image. Such samples are detected due to the fact that a profile with an outlier sum would contain a fragment which would not have been generated by restriction of the PCR product or would have been subsequently contaminated.

With the function RFLPlod, which can be used to remove all bands below a given threshold, the package deals with uncertain or dubious bands of shorter lengths which may not be recorded or cannot be sized in a fraction of the samples.

Subsequently, the RFLP samples are grouped by fragment numbers (function nrBands), based on the assumption that only samples with identical band numbers could belong to identical genotypes. RFLPtools then computes distances between the fragment patterns of RFLP samples within each of these groups (function RFLPdist), based on the molecular weight of the bands, evoking the function dist (base R package stats) with default Euclidean distance. Alternatively, different distance methods such as 'Manhattan' or 'Canberra' or even completely different functions can be used. The use of migration distances of the molecular weights of fragments instead of band matching is the main advantage of RFLPtools. As RFLPtools generates objects of basic R data types, high compatibility with other R libraries and analysis methods are ensured, which is considered a clear advantage over other existing software solutions.

In cases where band detection is uncertain, for example, due to density variation between gels and samples, a second method of similarity analysis can be used (function RFLPdist2). This involves the computation of the distance between the molecular weight of a sample S1 with x bands and a sample S2 with x + y bands and the distances between the molecular weight of sample S1 and the molecular weight of all possible combinations of S2 with x bands are computed. The distance between S1 and S2 is then defined as the minimum of these distances.

Let us consider an artificial example: sample S1 has two bands with molecular weights m1 and m2 (m1 > m2); sample S2 has three bands with molecular weights n1, n2 and n3 (n1 > n2 > n3). RFLPdist2 computes the distance

between S1 and (n1, n2), (n1, n3), (n2, n3), and the minimum of the three distances is returned. Let us further assume that m1 = n1 and m2 = n3 then the distance between S1 and (n1, n3) is zero, which is also the minimum distance returned by RFLPdist2. There is also an option to set a lower limit, LOD. If LOD is specified, it is assumed that missing bands occur only below this threshold, that is, the number of bands larger than or equal to LOD has to be identical and all possible combinations are only considered for bands smaller than the LOD.

Furthermore, slight variation between gels and resulting differences in molecular weight detection between identical RFLP types, for instance due to smiley effects, can be anticipated using the functions diffDist and linCombDist. The function diffDist computes and returns the distances between the rows of a data matrix, where instead of the row values as in the case of 'dist', the successive differences of the row values are used. The function might be helpful if there is a shift with respect to the measured bands, that is, the Euclidean distance of (550, 500, 300, 250) and (510, 460, 260, 210) will be 0 instead of 80. Additionally, the function linCombDist was implemented, which uses a linear combination of distances and provides a way to combine molecular weights and band spacing to calculate the similarity of samples, with the possibility of choosing weights for both methods. With the help of linCombDist, two distance measures can be simultaneously specified to compute the distances between the rows of a data matrix. Depending on the chosen weights w1 and w2, a linear (w1, w2 arbitrary) or convex (w1 + w2 = 1) combination of the two results can be calculated.

In case multiple data sets from a single sample (i.e. separate digests with two or more enzymes) are available, the function RFLPcombine can be used to combine an arbitrary number of data sets.

Based on the calculated distances, hierarchical cluster analysis of RFLP samples can be performed using the function hclust (R base package stats), where the default clustering method is 'complete linkage'. Other clustering or unsupervised learning methods such as multidimensional scaling can also be applied easily. To obtain the final groups of identical samples, the value at which the dendrogram should be cut (cutree, R base package stats)

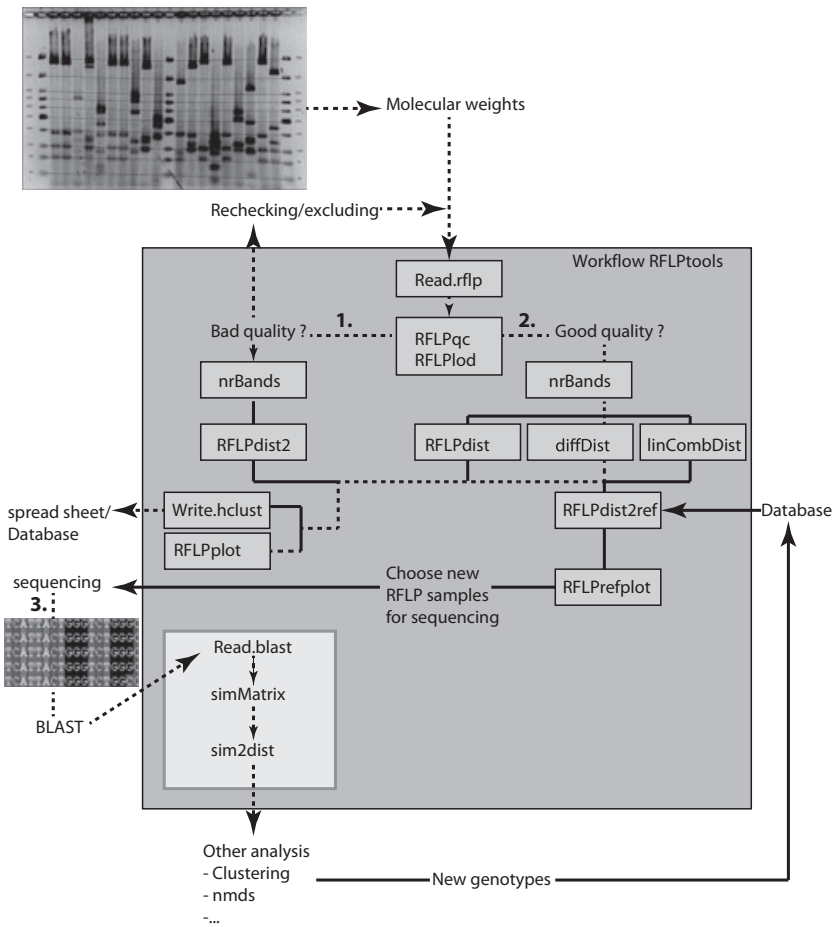

Fig. 1 Workflow of RFLP samples during diversity studies. The pale grey area summarizes functions included in R package RFLPtools. Dotted lines represent the workflow applied on the example data: 1. The RFLP input file contained manipulated samples, which were excluded after RFLPqc. 2. A new input file was analysed, an RFLPplot was drawn (see Fig. 2), and the results of clustering were exported using write.rflp. 3. Sequences of all RFLP samples in the example data were analysed.

must be considered by the scientist. Therefore, the similarity of samples (e.g. in terms of Euclidean distance) can be plotted using dendrograms and parallel molecular weight patterns, which enable a rapid and rather easy detection of identical samples (function RFLPplot) (see Fig. 1). After cutting the dendrograms (cutree, R base package stats), the function write.rflp creates an output file containing the cluster group assignment for each sample. This text file can easily be processed further, for example, in spreadsheet software applications.

During a biodiversity study using PCR-RFLP, the resulting RFLP groups must be verified by sequencing two or more samples from each cluster and checking their identity, to ensure that the cutting height and classification of fragment patterns were low enough so that the aggregation of different genotypes in one cluster is avoided. To facilitate the rapid identification of PCR-RFLP samples, two or more representative samples (PCR products) from each detected similarity group can be sequenced, and the resulting nucleotide sequences compared with data in the NCBI GenBank (http://www.ncbi.nlm.nih.gov) or other sequence databases. Reference data sets including band patterns and taxonomic affiliations might subsequently be established.

Common phylogenetic approaches require alignments of sequences. This is not a problem with highly similar sequences, but sequences deriving from environmental samples might include organisms of widely different taxa, which might be aligned with difficulty. We recommend an approach that includes pairwise similarities, calculated by the BLAST standalone tool, and enables the comparison of different taxa with no need for a detailed alignment. The success of this approach was shown in Peršoh *et al.* (2010), Peršoh & Rambold (2012), Flessa *et al.* (2012) and Flessa & Rambold (2013). RFLPtools enables the discrimination of DNA sequences via tabular report files of standalone BLAST (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release). Therefore, the similarity values for all-vs.-all BLAST results of DNA sequences generated with standalone BLAST from NCBI or local BLAST, as implemented in BioEdit v7.0.9 (Hall 1999), for example, are used to build a similarity matrix. The BLAST search tool is not implemented in RFLPtools, but tabular output files can easily be used for data exchange between the above-mentioned software programs. To import and analyse BLAST tabular report files, RFLPtools provides the function read.blast. Subsequently, the data frame obtained is modified into a similarity matrix. Details of the algorithm implemented to compute the similarity between samples can be looked up in the help files of the package. A visualization of the similarity matrix is possible using, for example, simPlot (R package MKmisc; Kohl 2012). As a terminal step, identified RFLP samples can be used to establish a reference data set. In the case of an existing reference data set containing RFLP fragment patterns and already identified taxon names (e.g. via DNA sequencing or BLAST search results), a comparison of RFLP genotypes derived from a new study with the existing reference samples is possible and enables the detection of known genotypes (function RFLPrefdist). The implemented method RFLPrefplot facilitates a visual comparison of new samples with reference samples (see Fig. 2). With these options, RFLPtools provides a PCR-RFLP-adapted application, which overcomes general problems when handling RFLP-derived data.

As we are generating objects of basic data types, the resulting compatibility with other R packages (i.e. population genetic packages) and analysis methods is a further advantage of RFLPtools. For example, it can be used with the popgen package, which uses R standard datatypes such as matrix, array and vector, or with the adegenet package, which contains several convert functions; that is, R standard datatypes can easily be converted to objects of adegenet classes.

Application example

To demonstrate the application advantages of RFLPtools, we selected a data set (see Appendix S1, Supporting information) including 112 RFLP types generated from double-stranded fungal ITS rRNA sequences obtained from cultivated fungi in previous studies (Triebel *et al.* 2005; Peršoh *et al.* 2010). Sequences were deposited at EMBL under the accession nos FR773168–FR773170, FR773172, FR773288, FR773289, FR773296, FR773297, FR773300, FR773304–FR773306, FR773308, FR773321, FR774049–FR774075, FR774077–FR774079, FR774081, FR774084, FR774089, FR774090, FR774092, FR774098, FR774101–FR774115, FR774117–FR774161.

Taxonomic names were assigned to the sequences using MegaBLAST (Zhang *et al.* 2000) at the NCBI website (http://www.ncbi.nlm.nih.gov; status: January 2010). A consensus name was compiled from the names under which sequences obtaining a 'bitscore' of at least 90% of the best matching sequences were deposited, following the approach of Peršoh *et al.* (2010). The nomenclature and classification concepts applied follow Index Fungorum (http://www.indexfungorum.org) and Myconet (http://www.fieldmuseum.org/myconet).

Restriction fragment length polymorphism band patterns were artificially created by cutting the ITS sequences with *Alu*I and *Msp*I in a virtual digest with the online application 'RestrictionMapper' (http://www.restrictionmapper.org/) and were stored in a text format suitable for the read.blast function, containing the sample name, number of bands and molecular weight of the band. The RFLP patterns of four sequences (FR773320, FR773324, FR774080 and 774082) were manually manipulated to

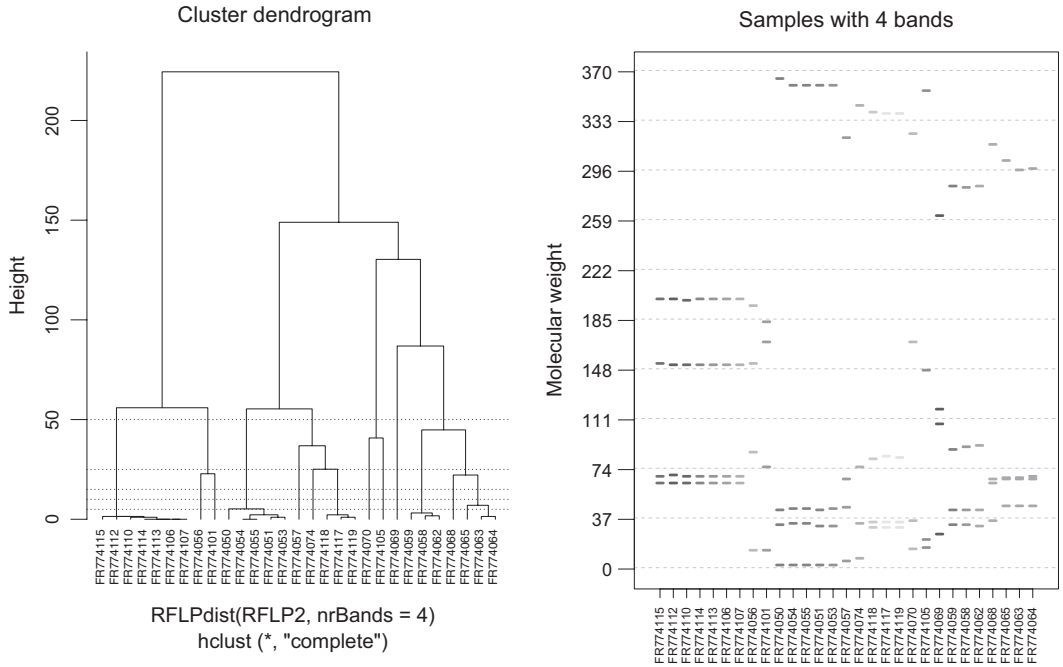

Fig. 2 Visual comparison of molecular weights from an example data set containing 29 RFLP samples exhibiting four fragments.

generate 'problematic' band patterns (in FR773320 and FR773324, the first band was deleted to simulate an overseen band, and in FR774070b and 774049b, which were manipulated duplicates from samples FR774070 and FR774049, one band was duplicated and modified with a difference in the molecular weight by three base pairs, to simulate an overexpressed thick band, which was identified as two bands by gel analysis software).

Subsequently, the RFLP samples were grouped by fragment numbers (function nrBands), resulting in eight groups containing RFLP samples with 1–8 bands, respectively. By applying the quality check function (RFLPqc, QC.lo = 0.9, QC.up = 1.1) on the data set, the four manipulated RFLP samples fell outside the range (the sum of bands of sample FR773320 was out of range by 40.57% and that of FR773324 by 40.58%, FR774080 by 131.07% and FR774082 by 127.29%) and were excluded from further analysis.

Distances between fragment patterns of RFLP samples within each of these groups were computed with the function RFLPdist. Subsequently, hierarchical cluster analysis of RFLP samples was performed using the function hclust (R base package stats). To extract groups of identical samples, the resultant dendrograms were cut (cutree, R base package stats) at heights of 5, 10, 15 and 25 to compare the resulting RFLP clusters and identify the optimal cutting height for these data. In our example, the resulting RFLP cluster can easily be compared with the genotype affiliation of each sample (for results, see Appendix S2, Supporting information). Therefore, the similarity of samples was plotted using dendrograms and parallel molecular weight patterns (function RFLPplot) (see Fig. 2). After cutting the dendrograms, the function write.hclust was applied to write an output file containing the cluster group assignment for each sample [included in Appendix S2, Supporting information (cluster number)].

This procedure results in 51 different RFLP groups at a height of 5. This height appears optimal for the data set used, because there is no RFLP cluster with different fungal genotypes at a height of $h = 5$. However, some of the 33 fungal genotypes are split into different RFLP clusters [*Aureobasidium pullulans*-1 (3 cluster), *Botryosphaeria*-1 (2 cluster), *Capnocheirides*-1 (2 cluster), *Cladosporium*-1 (3 cluster), *Cladosporium*-2 (5 cluster), *Phialocephala*-1 (2 cluster), *Fusarium*-1 (2 cluster), *Lewia*-1 (2 cluster), *and Vibrisseaceae*-1 (2 cluster)], but all of these clusters are homogeneous. At a height of $h = 10$, one cluster is nonhomogeneous (*Cladosporium*-1 and *Cladosporium*-2 are in the same group). In this data set, with $h = 5$, no fungal genotype was overseen, but there is a requirement for sequence identification of each RFLP cluster to detect genotypes divided into different RFLP types, due to mutations in the sequence. Identification of split groups is important to avoid an overestimation of the fungal diversity. If RFLPtools is used as a presort tool in diversity studies, there is no need to sequence each individual RFLP sample.

Conclusion

Polymerase chain reaction-restriction fragment length polymorphism is a common and applicable method in

diversity studies, but a suitable, freely available analysis application that provides interoperability, flexibility and compatibility with further statistical analysis methods is still needed. Although several software applications exist for PCR-RFLP fragment pattern analysis, RFLPtools represents the first purpose-built application for PCR-RFLP in R. R packages have the advantage that statistical analyses are easily applicable to the resulting outputs, and as objects of basic data types are generated, the resulting compatibility with other R packages and analysis methods are a further advantage of RFLPtools.

Because its function composition follows the workflow of PCR-RFLP analysis, RFLPtools allows reliable data processing from the fragment band pattern to the identification of RFLP samples. Furthermore, in contrast to existing applications, RFLPtools provides a quality check for the input data. This option for data quality checking and band pattern visualization supports the handling of problematic data by detecting invalid samples and allows the subsequent rechecking or exclusion of those samples, which improves data quality. As RFLPtools was designed for data gathered from images of electrophoresis gels, the package contains some functions to handle biased data. However, the additional functions might be applied to data gathered from capillary sequencers using fluorescently labelled primers.

The main advantage of RFLPtools is that it uses distances of the molecular weights of fragments instead of band matching. Options to generate a reference data set of identified RFLP samples based on DNA sequences and to report files derived from all-vs.-all BLAST searches allow for the comparison of new samples with known genotypes, supplementing the application. This is completed by a second option for visualization of band patterns of new samples and references. With these options, RFLPtools provides a PCR-RFLP-adapted application which considers general problems encountered when handling RFLP-derived data.

Acknowledgements

The authors thank Derek Peršoh for valuable comments on parts of the function RFLPqc and simMatrix Gerhard Rambold who acted supportive through the development of RFLPtools package and Nicole Schwenger for additional testing with biological data sets. In addition, we would like to thank the anonymous reviewers and the managing editor for valuable comments which helped us to improve our package RFLPtools.

References

Dickie IA, Avis PG, McLaughlin DJ, Reich PB (2003) Good-Enough RFLP Matcher (GERM) program. *Mycorrhiza*, **13**, 171–172.

Duran A, Slippers B, Gryzenhout M *et al*. (2009) DNA-based method for rapid identification of the pine pathogen, *Phytophthora pinifolia*. *FEMS Microbiology Letters*, **298**, 99–104.

Flessa F, Rambold G (2013) *Diversity of the Capnocheirides Rhododendri-Dominated Fungal Community in the Phyllosphere of Rhododendron Ferrugineum L*. Nova Hedwigia accepted.

Flessa F, Rambold G, Peršoh D (2012) Annuality of Central European deciduous tree leaves delimits community development of epifoliar pigmented fungi. *Fungal Ecology*, **5**, 554–561.

Goldstein DB, Schlötterer C (ed.) (1999) *Microsatellites – Evolution and Applications*. Oxford University Press, Oxford, UK.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Kohl M (2012) *MKmisc: Miscellaneous Functions From M. Kohl*. R package version 0.92. R, Vienna, Austria.

Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology*, **63**, 4516–4522.

Peršoh D, Rambold G (2012) Lichen-associated fungi of the *Letharietum vulpinae*. *Mycological Progress*, **11**, 753–760.

Peršoh D, Melcher M, Flessa F, Rambold G (2010) First fungal community analyses of endophytic ascomycetes associated with *Viscum album* ssp. *austriacum* and its host *Pinus sylvestris*. *Fungal Biology*, **114**, 585–596.

R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Saari TA, Saari SK, Campbell CD, Alexander IJ, Anderson IC (2007) FragMatch – a program for the analysis of DNA fragment data. *Mycorrhiza*, **17**, 133–136.

Sykorova Z, Wiemken A, Redecker D (2007) Cooccurring *Gentiana verna* and *Gentiana acaulis* and their neighboring plants in two Swiss upper montane meadows harbor distinct arbuscular mycorrhizal fungal communities. *Applied and Environmental Microbiology*, **73**, 5426–5434.

Triebel D, Peršoh D, Wollweber H, Stadler M (2005) Phylogenetic relationships among *Daldinia, Entonaema*, and *Hypoxylon* as inferred from ITS nrDNA analyses of Xylariales. *Nova Hedwigia*, **80**, 25–43.

Tsarouhas V, Gullberg U, Lagercrantz U (2002) An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in Salix. *Theoretical and Applied Genetics*, **105**, 277–288.

Vos P, Hogers R, Bleeker M *et al*. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Watrud LS, Martin K, Donegan KK, Stone JK, Coleman CG (2006) Comparison of taxonomic, colony morphotype and PCR-RELP methods to characterize microfungal diversity. *Mycologia*, **98**, 384–392.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

A.K. and F.F. designed the research; M.K. programmed and described the functions in the manuscript, F.F. performed the analysis and wrote the description of the application example, A.K. and F.F. wrote the paper.

Data Accessibility

The software package is freely available from 'The Comprehensive R Archive Network' (CRAN) at http://cran.r-project.org/web/packages/RFLPtools.

DNA sequences: GenBank accessions FR773168–FR773170, FR773172, FR773288, FR773289, FR773296,

FR773297, FR773300, FR773304–FR773306, FR773308, FR773321, FR774049–FR774075, FR774077–FR774079, FR774081, FR774084, FR774089, FR774090, FR774092, FR774098, FR774101–FR774115, FR774117–FR774161.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Example dataset.

Appendix S2 Comparison of four different cutting heights of resultant dendrograms by applying cutree on the example dataset (section application example).