J. Phytopathology **146**, 427–430 (1998)

ISSN 0931-1785

USDA-ARS, Foreign Disease-Weed Science Research Unit, Frederick, MD USA

Phylogenetic Inference Based on Information Theory-Based PCR Amplification

P. W. TOOLEY[1], J. J. SALVO[2], T. D. SCHNEIDER[3] and P. K. ROGAN[4]

[1]USDA-ARS, 1301 Ditto Ave., Fort Detrick, MD 21702–5023, USA; [2]General Electric Co., PO Box 8, Schenectady, NY 12301, USA; [3]NCI-FCRDC, Laboratory of Experimental and Computational Biology, PO Box B, Frederick, MD 21702–1201, USA; [4]Department of Human Genetics, Allegheny University of the Health Sciences, 320 E. North Ave., Pittsburgh, PA 15212–4772, USA (correspondence to P. W. Tooley)

With 2 figures

Received December 8, 1997; accepted February 19, 1998

Abstract

As a method for the determination of the taxonomic affinities of plant pathogens and other organisms, a set of 'universal' polymerase chain reaction (PCR) primers which amplify a taxonomically diverse sequence domain of 28S ribosomal DNA (rDNA) were designed. The PCR primers chosen by information-theory analysis generated PCR products using DNA templates from a wide diversity of organisms. Sequences of PCR products were then obtained which allowed phylogenetic dendrograms to be constructed. Based on the above analysis, the Oomycete pathogen *Phytophthora infestans* clustered with the protist *Prorocentrum micans* rather than with representatives of the true fungi, consistent with its designation as a 'pseudofungus'. *Magnaporthe grisea*, another important plant pathogen, clustered with the true fungi as expected. The approach described can be used with other plant pathogens to clarify phylogeny of new or ambiguously designated species.

Zusammenfassung

Phylogenetische Erkenntnisse auf Grundlage einer auf Informationstheorie basierenden PCR-Amplifikation

Als Verfahren zur Bestimmung taxonomischer Ähnlichkeiten von Pflanzenpathogenen und anderen Organismen entwickelten wir einen Satz 'universeller' PCR-Primer, die einen bei verschiedenen Taxa unterschiedlichen Sequenzbereich von 28S-ribosomaler DNA (rDNA) amplifizieren. Mit Hilfe der informationstheoretischen Analyse ausgewählte PCR-Primer erzeugten PCR-Produkte von DNA-Matrizen ganz unterschiedlicher Organismen. Wir erhielten Sequenzen von PCR-Produkten, die eine Konstruktion phylogenetischer Dendrogramme erlaubten. Basierend auf der oben genannten Analyse clusterte das zu den Oomyceten gehörende Pathogen *Phytophthora infestans* mit dem Einzeller *Prorocentrum micans* und nicht mit höheren Pilzen, was mit seiner Einstufung als 'Scheinpilz' übereinstimmt. *Magnaporthe grisea*, ein anderes wichtiges Pflanzenpathogen, clusterte dagegen erwartungsgemäß mit den höheren Pilzen. Das von uns beschriebene Verfahren kann auch für andere Pflanzenpathogene angewendet werden, um die Phylogenese neuer Arten zu untersuchen oder bei strittiger Nomenklatur Klärung herbeizuführen.

Introduction

The use of DNA sequences to classify organisms and derive phylogenies has become widespread (Fitch and Margoliash, 1967; Woese, 1987; Field et al., 1988; Sogin, 1990; Embly et al., 1994). However, problems exist with certain regions of DNA, such as 5S ribosomal DNA in which too few variable sites may exist for meaningful analysis. Such uninformative regions of DNA cannot thus be used to identify unknown organisms. To add information which may help clarify the taxonomic status of the pathogen *Phytophthora infestans*, the method of visual display of sequence conservation has been applied in identifying a region of sequence divergence flanked by two regions of conservation. The degree of conservation of the sequence is visualized with a sequence 'logo' (Schneider and Stephens, 1990), which displays the nucleotides graphically along with their information (sequence conservation) content. Previously, information analysis was performed on a set of aligned 28S ribosomal DNA (rDNA) sequences from five phyla to select the region to be amplified (Rogan et al., 1995). The sequence chosen was 28S rDNA, for which the 5′ and 3′ terminal coordinates of the PCR product correspond to positions 2698 and 2849 of the human sequence (GenBank entry HUMRGM, accession number M11167) (Rogan et al., 1995). The variable domain chosen is readable on a single sequencing gel so that rapid, low-resolution taxonomic designations can be made. The benefit of this approach is to satisfy the requirement of PCR for two conserved sequences from which to perform DNA amplification, and the desirability from a taxonomic perspective of

amplifying a highly variable region from which to make evolutionary comparisons. Thus, the sequences of ribosomal RNAs from widely divergent species can be aligned and phylogenetic relationships assessed. This approach lends itself to the identification of unknown organisms by placing them on a phylogenetic tree.

Previously, the above primers were shown to amplify the variable domain in 28S rDNA in a wide variety of eukaryotes (Rogan et al., 1995). The purpose of the present studies was to obtain sequence information of this region for the ambiguously classified *P. infestans* as well as the taxonomically well-defined rice blast fungus *Magnaporthe grisea*, and determine the utility of this approach in the placement of these organisms on a phylogenetic tree. The general utility of the information theory approach in its application to diverse organisms and the classification of such organisms through phylogenetic analysis has also been determined.

Materials and Methods

DNA was extracted from fungal plant pathogens *M. grisea* and *P. infestans* using the method described by Goodwin et al. (1992). Human placental DNA was extracted using standard protocols (Sambrook et al., 1989). DNA of *Saccharomyces cerevisiae* and *Pichia pinus* were obtained from the laboratory of Dr Jeffrey Strathern, National Cancer Institute, Frederick, MD, USA.

Full-length 28S ribosomal DNA sequences representing a broad taxonomic distribution and obtained from the GenBank repository (Burks et al., 1991) were aligned using a rectification algorithm (Feng and Doolittle, 1987; Higgins and Sharp, 1989) in which the human sequence was chosen as a reference. A sequence logo (Schneider and Stephens, 1990) was created from the aligned 28S sequences, and a region selected with two conserved regions (ones with greater than 1.5 bits per position) flanking a divergent region (with less than 0.5 bits per position) (Fig. 1). The sequence logo was used to choose two PCR primers which were in regions of high conservation (in bits) but flanking a single contiguous region of low conservation of information content. Sequences at the 3′ termini of the primers consisted of invariant segments (>3 nucleotides), so the primer end which is extended by the DNA polymerase is consistently annealed to the DNA. The primers were also designed to contain restriction sites useful for subsequent cloning of the amplification products. The amplified human PCR product was predicted to be 159 base pairs long (positions 10–168 of GenBank entry HUMRGM, accession number M11167) and the primers were as follows: *Sac*I primer: 5′-GGTGAGCTCTCGCTGGCCCTTGA-3′, and *Bam*HI primer: 5′-GTTACGGATCCGGCTTGCCGACTTC-3′. The PCR cycling conditions consisted of 4 min at 94°C followed by 40 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min followed by a final step at 72°C for 7 min.

The DNA was cloned into the m13 mp19 vector after digestion of the vector with *Sac*I, *BamH*I, and *Kpn*I restriction enzymes. The DNA hybridization analysis was carried out by running 1–3 μg of DNA on 0.7% agarose gels following restriction with *Hin*DIII. The DNA fragments were denatured and blotted onto Nylon (Micron Separations Inc., Westborough, MA, USA) membranes by capillary transfer for 16 h. Prehybridization was carried out at 65°C for 1 h in 0.25 M $NaHPO_4$ (pH 7.2) – 0.25 M NaCl-7% sodium dodecyl sulphate (SDS) – 1 mM EDTA (Sigma Chemical Co., St. Louis, MI, USA) (Amasino, 1986). Hybridization was performed at the same temperature for 14–16 h after the addition of radiolabelled probe. The membranes were washed at 65°C for 20 min in 2 X SSC – 0.1% SDS and twice in a 0.1 X SSC – 0.1% SDS solutionand then they were exposed to X-ray film at −80°C for 24–72 h.

Single stranded DNA was produced using a standard protocol (Amersham, Arligton Heights, IL, USA) and DNA sequenced manually using Sequenase version 2.0 and 6% acrylamide gels. DNA sequences were analysed using the GCG software package on a VAX mainframe computer at the Frederick Biomedical Supercomputing

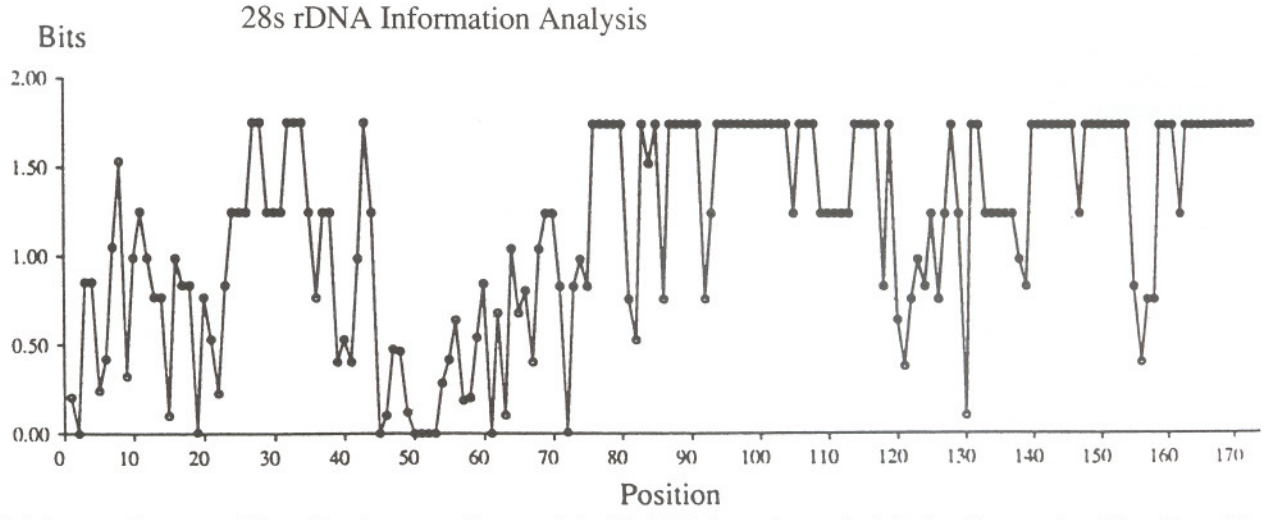

Fig. 1 Sequence logo created from aligned sequences for part of the 28S rDNA from nine species including *Homo sapiens*, *Citrus limon*, *Mucor racemosis*, *Oryza sativa*, *Physarum polycephalum*, *Lycopersicon esculentum*, *Caenorhabditis elegans*, *Xenopus laevis*, and *Saccharomyces cerevisiae* (Rogan et al., 1995)

Facility (Devereaux et al., 1984). Sequences were analysed in PHYLIP for DNA distance analyses and bootstrap analysis (Felsenstein, 1993).

Results

The PCR amplification products of the expected size (159 bp) obtained for all organisms tested were cloned into the vector m13 mp19 and used as probes on southern blots of restriction digested genomic DNA to confirm that amplification products hybridized with rDNA of the organisms. The PCR products showed hybridization to multicopy conserved restriction fragments, consistent with tandem repeat organization (data not shown). Sequences of the amplified 28S rDNA region were aligned with sequences of the same region derived from Genbank for a group of organisms which included *Xenopus laevis*, *Prorocentrum micans*, *Caenorhabditis elegans*, *Lycopersicon esculentum*, *Oryza sativa*, *Citrus limon*, *Mucor racemosus*, and *Physarum polycephalum*. Five of the sequences obtained were combined with the eight GenBank sequences in a data set for which DNA distance values were obtained using PHYLIP. The data were then subjected to bootstrap analysis using SEQBOOT within PHYLIP and a dendrogram produced showing phylogenetic relationships among the organisms based on the chosen 28S rDNA region (Fig. 2). The three plant species represented (*L. esculentum*, *O. sativa*, and *C. limon*) all grouped together, as did the four species of true fungi (plant pathogen *M. grisea*, *P. pinus*, *S. cerevisiae*, and *M. racemosus*). *Phytophthora infestans* grouped with the protist *P. micans* and the nematode *C. elegans*. The slime mould *P. polycephalum*, the amphibian *X. laevis*, and *Homo sapiens* all were placed distantly from one another and from the other organisms in the analysis (Fig. 2).

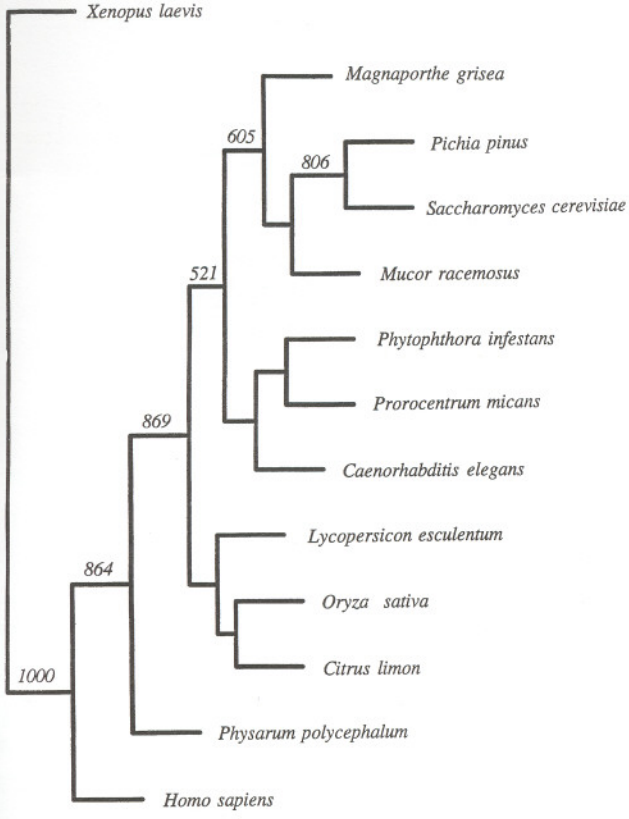

Fig. 2 Consensus tree derived from bootstrap analysis in PHYLIP of 28S rDNA sequences from 13 different organisms amplified using primers developed based on information theory. Bootstrap values are shown on the branches (values less than 500 were omitted)

Discussion

In this study, the application of a general approach to comparing organisms based on information analysis of a known set of 28S rDNA sequences, which was used to select a phylogenetically informative domain has been demonstrated. Although a relatively short sequence was analysed, the dendrogram derived from these sequences groups the organisms into phylogenetically reasonable groups. The destructive plant pathogen *P. infestans* grouped with the protist *P. micans* and the nematode *C. elegans*, and not with the true fungi included in the study. This finding lends support to the hypothesis that, evolutionarily, *P. infestans* is thought to be more closely related to heterokont algae and other members of the 'fifth kingdom' (Margulis and Schwartz, 1982) than to the true fungi due to features such as the heterokont pattern of zoospore flagellation, mitochondria with tubular cristae, and other characters (Margulis and Schwartz, 1982; Cavalier-Smith, 1987; Beakes, 1989; Margulis et al., 1990).

In addition to clarifying taxonomic relationships among plant pathogens, other potential applications of the method described include rapidly generating 28S rDNA sequences from different organisms (e.g. polymorphism studies of intergenic spacer regions), resolving ambiguous taxononic designations and classifying unknown or anonymous specimens (i.e. where morphological characters can no longer be assessed). With the accumulating data set for the selected sequence being generated from known species, it may be possible to use this approach to make approximate identifications of unknown, unclassified, or damaged specimens.

Acknowledgements

The authors would like to thank Britt Bunyard for assistance with the PHYLIP computer program used in these studies.

Literature

Amasino, R. M. (1986): Acceleration of nucleic acid hybridization rate by polyethylene glycol. Anal. Biochem. **152,** 304–307.

Beakes, G. W. (1989): Oomycete fungi: their phylogeny and relationship to chromophyte algae. In: Green, J. C., B. S. C. Leadbeater and W. L. Diver (eds), The Chromophyte Algae: Problems and Perspectives, pp. 325–355. Clarendon Press, Oxford.

Burks, C., M. Cassidy, M. J. Cinkosky, K. E. Cumella, P. Gilna, J. E.-H. Hayden, G. M. Keen, T. A. Kelly, M. Kelly, D. Kristofferson, J. Ryals (1991): 'GenBank'. Nucl. Acids Res. **19,** 2221–2225.

Cavalier-Smith, T. (1987): The origin of fungi and pseudofungi. In: Rayner, A. D. M., C. M. Brasier and D. Moore, (eds), Evolutionary Biology of the Fungi, pp. 339–353, Cambridge University Press, Cambridge.

Devereaux, J., P. Haeberli, O. Smithies (1984): A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12,** 387–395.

Embley, T. M., R. P. Hirt, D. M. Williams (1994): Biodiversity at the molecular level: the domains, kingdoms and phyla of life. Phil. Trans. R. Soc. London B. **345,** 21–33.

Felsenstein, J. (1993): PHYLIP: Phylogeny Inference Package User's Guide. Frederick Biomedical Supercomputing Center, Frederick, MD.

Feng, D. F., R. F. Doolittle (1987): Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25,** 351–360.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace, R. A. Raff (1988): Molecular phylogeny of the animal kingdom. Science **239,** 748–753.

Fitch, W. M., E. Margoliash (1967): Construction of phylogenetic trees. Science **155,** 279–284.

Goodwin, S. B., A. Drenth, W. E. Fry (1992): Cloning and genetic analyses of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. **22,** 107–115.

Higgins, D. G., P. M. Sharp (1989): Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS **5,** 151–153.

Margulis, L., J. O. Corliss, M. Melkonian, D. J. Chapman (eds) (1990): Handbook of Protoctista. Jones and Bartlett, Boston, MA.

Margulis, L., K. V. Schwartz (1982): Five Kingdoms: an Illustrated guide to the Phyla of Life on Earth. W. H. Freeman, San Francisco.

Rogan, P. K., J. J. Salvo, R. M. Stephens, T. D. Schneider (1995): Visual display of sequence conservation as an aid to taxonomic classification using PCR amplification. In: Pickover, C. A. (ed.), Visualizing Biological Information, pp. 21–32, World Scientific, River Edge, NJ.

Sambrook, J., E. F. Fritsch, T. Maniatis (1989): Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schneider, T. D., R. M. Stephens (1990): Sequence logos: a new way to display consensus sequences. Nucl. Acids Res. **18,** 6097–6100.

Sogin, M. L. (1990): Amplification of ribosomal RNA genes for molecular evolution studies. In: Innis, M. A., D. H. Gelfand, J. J. Sninsky and T. J. White (ed.), PCR Protocols, pp. 307–314. Academic Press, San Diego, CA.

Woese, C. R. (1987): Bacterial evolution. Microbiol. Rev. **51,** 221–271.