Gene, 109 (1991) 161–165
© 1991 Elsevier Science Publishers B.V. All rights reserved. 0378-1119/91/$03.50

GENE 06189

The *TRP1* gene of *Phytophthora parasitica* encoding indole-3-glycerolphosphate synthase–*N*-(5′-phosphoribosyl)anthranilate isomerase: structure and evolutionary distance from homologous fungal genes

(Oomycetes; tryptophan biosynthesis; G + C content; introns; Ω interposon; phylogenetic tree)

Petr Karlovsky and Hermann H. Prell

Institut für Pflanzenpathologie und Pflanzenschutz der Universität Göttingen, and (formerly) Abteilung für Molekulare Genetik der Gesellschaft für Strahlen- und Umweltforschung mbH, Munich (F.R.G.)

Received by H.M. Krisch: 18 April 1991
Revised/Accepted: 1 August/11 September 1991
Received at publishers: 23 September 1991

SUMMARY

The *TRP1* gene was isolated from the genome of *Phytophthora parasitica*. It encodes bifunctional enzyme of the tryptophan biosynthetic pathway indole-3-glycerolphosphate synthase–*N*-(5′-phosphoribosyl)anthranilate isomerase (IGPS–PRAI). The gene was localized and sequenced using random in vitro insertions of Ω interposon. The domain structure of the protein product was found to be similar to that of enteric bacteria but different from the structure of homologous enzymes in fungi. Two introns in the IGPS domain were found. This is unique in eukaryotic IGPS-encoding genes so far sequenced. Comparative analysis of the primary structure of IGPS and PRAI domains [neighbor-joining method of Saitou and Nei, Mol. Biol. Evol. 44 (1987) 406–425] confirmed a large phylogenetic distance of *TRP1* from corresponding fungal genes. In the resulting distance tree *Phytophthora* sequences are located outside of the cluster which encompasses all known homologous proteins from fungi indicating that the lineage of oomycetes took a separate course of development before speciation within the fungal line of descent began. Two of the oligopeptide insertions engineered into the F domain of the protein product did not abolish the enzymatic activity of the protein.

INTRODUCTION

Tryptophan biosynthesis belongs to the most detailed studied metabolic pathways on the genetic level. Genes encoding IGPS (domain C in classical terminology) and PRAI (domain F) were isolated from many bacterial species (Crawford, 1989), from one archaebacterium, and from several eukaryotes (Hütter and Niederberger, 1986). Genetic determinants responsible for both enzymatic activities are organized in distinctly different ways in various organisms. In filamentous fungi, these domains together with the G-domain (glutamine amidotransferase) constitute a trifunctional protein G-C-F, while in yeast they are separated into the isolated domain F and a bifunctional G-C protein. Still another organization comprising a C-F fusion similar to that occurring in enteric bacteria was suggested to exist in oomycetes and in blue-green algae on the basis of sedimentation sucrose gradient analysis of enzyme activities (Hütter and DeMoss, 1967). However, no data concerning tryptophan gene structure were available in these phylogenetic groups so far.

Correspondence to: Dr. P. Karlovsky, Institute of Plant Pathology, Grisebachstrasse 6, W-3400 Göttingen (F.R.G.)
Tel. (49-551)399580; Fax (49-551)394187.

Abbreviations: aa, amino acid(s); bp, base pair(s); dNTP, deoxyribonucleoside triphosphate; IGPS, indole-3-glycerolphosphate synthase; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); *P.*, *Phytophthora*; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; PRAI, *N*-(5′-phosphoribosyl)anthranilate isomerase; Ω, interposon (see section a).

EXPERIMENTAL AND DISCUSSION

(a) Isolation and insertion mutagenesis of *TRP1*

A gene library of nuclear DNA of *P. parasitica* (Fig. 1 legend) was transformed into *E. coli* strains carrying 13 different auxotrophic mutations and transformants were assayed on minimal media for complementation. Plasmids complementing mutations *hisB*463 and *trpC*1117 (defective in PRAI) were identified but no complementation was achieved for *argI*, *argHI*, *aroD6*, *leuB6*, *lysA*, *proA*, *purA*, *pyrE*, *trpA*33, *trpB* and *trpC*670 mutations. One of the *trpC*1117 complementing plasmids (pTR1) was chosen for further study and the *Phytophthora* gene responsible for the complementation was named *TRP1*. Introducing *lacI*^{Q} allele into the complemented strain (results not shown) and subcloning (Fig. 1) revealed that the *trpC*1117 complementing function is controlled by a *Phytophthora* sequence functioning as promoter in *E. coli*. The plasmid did not complement an IGPS mutation in *E. coli* (results not shown).

We used insertion mutagenesis in vitro at random sites employing the *Ω* interposon (Prentki and Krisch, 1984) (*i*) to localize the complementing sequence, (*ii*) to provide starting points for sequencing, and (*iii*) to introduce restriction sites for separating regulatory sequences and using them in the construction of transformation vectors for *Phytophthora*. A total number of 86 insertions were analyzed, 62 of which mapped within the cloned *Phytophthora* fragment. After deleting most of the *Ω* fragment with *Bam*HI about two thirds of the plasmids regained the ability to complement the tryptophan requirement of *trpC*1117. Physical mapping of the insertions delimited the region involved in complementation of PRAI to 820 bp (Fig. 1).

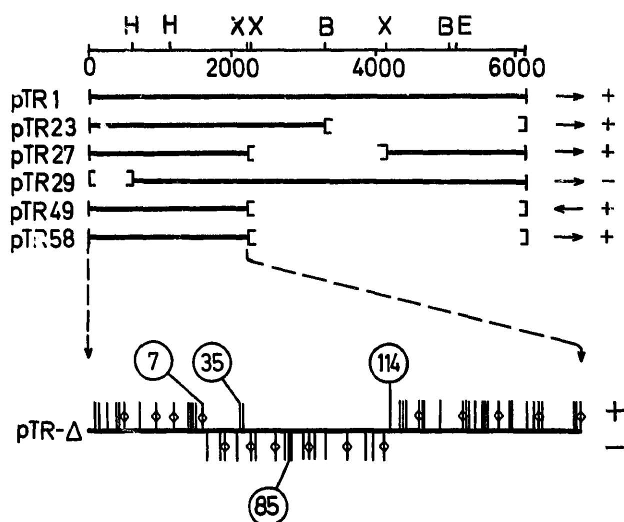

Fig. 1. Restriction map and *Ω* mutagenesis of the cloned *TRP1* gene. A map of the *Phytophthora* DNA fragment cloned in pTR1 is shown at the top, the scale is in bp. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; X, *Xho*I. Deletion derivatives of pTR1 are shown below it. Arrows on the right margin indicate the direction of transcription from the *lac* promoter through the fragment, plus and minus symbols stand for the ability to complement *trpC*1117 mutation in *E. coli*. The lower part of the figure is an expanded view of the insert in pTR58 with the location of *Ω* insertions. Derivatives which retain the complementary activity after the removal of *Ω* by *Bam*HI digestion and recircularization are drawn as short vertical lines above the axis, while derivatives manifesting abolished complementation are drawn as short lines under the axis. Insertions mentioned in the text are identified by numbers, insertions used for sequencing are indicated by diamonds. **Construction of the genome library:** Nuclear DNA of *P. parasitica* was partially digested with *Sau*3A and separated on a sucrose gradient. Fragments ranging from 7 to 12 kb were cloned in pUC19. ***Ω* mutagenesis:** Five μg of DNaseI-linearized PolIk-treated pTR58 were ligated with 2 μg of *Ω* in 500 μl of 25 mM Tris·HCl pH 7.5/5 mM $MgCl_2$/0.5 mM dithiothreitol/0.6 mM ATP/1 mM hexammine cobalt(III) chloride/2 units T4 DNA ligase for 30 min at 24°C. This ligation generated only linear products. They were separated on a gel and a band of fragments of the size of the sum of pTR58 and *Ω* was cut out. DNA was isolated from the gel and recircularized in 50 μl of the buffer as above but with only 0.3 mM ATP and with 6% polyethylene glycol 4000 (Sigma) instead of hexamine cobalt. The ligation mixture was used to transform *E. coli recA*$^-$, transformants were selected on plates with ampicillin (50 μg/ml) and spectinomycin (100 μg/ml).

(b) Nucleotide sequence of *TRP1*

The gene and its surrounding region were sequenced (3074 bp; acc. No. M64473) using *Bam*HI sites created by *Ω* insertions in pTR58 and BAL 31 deletions in pTR1 (Fig. 1). The *trpC*1117 complementing region was confirmed as PRAI coding sequence. The coding region of IGPS, interrupted by two putative introns, was found farther upstream. Thus the *TRP1* gene encodes a bifunctional protein with IGPS (domain C) and PRAI (domain F) activities (Fig. 2). Location of the first PRAI-inactivating

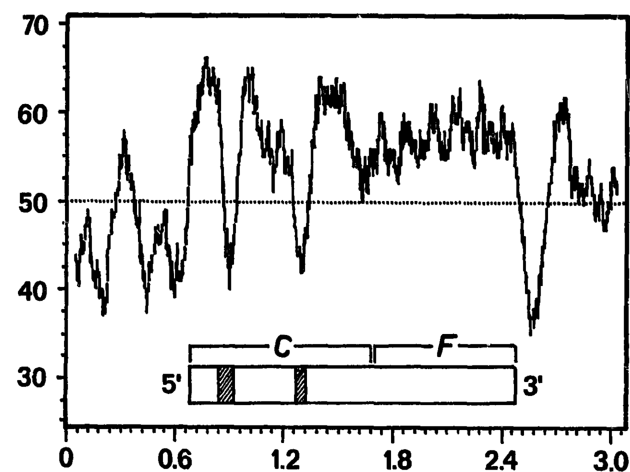

Fig. 2. Correlation between GC distribution and exon location in *TRP1*. Both strands were sequenced by chemical method (Maniatis et al., 1982) using *Bam*HI sites introduced by *Ω* mutagenesis in pTR58 (see diamonds in Fig. 1) and nested BAL 31 deletions made in pTR1 and its subclones. This sequence (3074 bp) is accessible in EMBL and GenBank databases under No. M64473. G + C content in % was computed using a window of 100 bp. The scale on the horizontal axis is in kb. Hatched boxes indicate introns within the coding sequence for the C domain.

insertion in the sequence (Fig. 1) suggests that translation in *E. coli* starts at Met^{267} (nt 1625) at the end of the C domain, far beyond both introns. Expression of *PRAI* but not of *IGPS* in *E. coli* was described similarly with *TRP1* genes of *Aspergillus* (Kos et al., 1988), *Phanerochaete* (Schrank et al., 1991) and *Neurospora* (Schechtman and Yanofsky, 1983).

Finding introns in the C domain of *TRP1* was rather surprising since introns had not been identified in any of the eukaryotic *IGPS* genes so far sequenced. Both introns have boundaries similar to the consensus 5′-GTRNGT...YAG typical of filamentous fungi, contain stop codons in phase with the IGPS reading frame, and possess an internal sequence CTAACT (nt 885 and 1286) matching the element believed to be involved in lariat formation during splicing.

The G + C content of the nuclear genome of *P. parasitica* is 49.0%–50.5% (Storck and Alexopoulos, 1970; Karlovsky and De Cock, 1991). As shown in the plot in Fig. 2, the G+C content is above 50% throughout the exons (58.3% on average) but it drops to low values around the predicted start and stop codons and in the putative introns (40.1% G+C on average). In fact, the plot could be used to precisely predict exon location using the cutoff value of 50% G+C.

A CATAAA sequence at nt position −89 from the start codon differs only in the least important nt from the TATA consensus (Bucher and Trifonov, 1986). At 22 bp downstream, a CA doublet followed mostly by pyrimidines marks a putative *tsp*. A CAAT box was not sought for since its existence was disproved statistically (Bucher and Trifonov, 1986).

A polyadenylation signal ACTAAA (consensus AATAAA) was found 30 bp downstream from the *TRP1*. A tract of alternating adenine-pyrimidine residues 5′-ACACACACATATA was located between the stop codon and the polyadenylation signal. A nonstandard helix geometry for this DNA segment was deduced by computer analysis (data not shown). Similar tracts of alternating purines and pyrimidines were found downstream from the *TRP1* gene of *Kluyveromyces lactis* (Stark and Milner, 1989) but their possible role in transcription regulation is unknown.

(c) Analysis of the putative aa sequence of the *TRP1* product and of its insertion mutations

The bifunctional organization of the *TRP1* product in *Phytophthora* confirms an almost quarter-century-old prediction based on the separation of enzyme activities in sucrose gradients which related oomycetes to algae rather than to fungi (Hütter and deMoss, 1967). Other known cases of such domain organization are documented only for enteric bacteria and for *Brevibacterium* (Crawford, 1989).

The C domain of *Phytophthora* is unusual in being disrupted by two introns. No introns were found in any other *IGPS* gene so far investigated. The location of the introns with respect to the three-dimensional structure of the protein obtained by aligning with the *E. coli* enzyme (data not shown) and the absence of introns in all other known eukaryotic *IGPS* genes supports the assumption that these introns arose through insertion into the *IGPS* gene after an ancestor of oomycetes diverged from the fungal lineage.

TABLE I

Similarity between IGPS and PRAI sequences of *Phytophthora parasitica* and related fungal sequences[a]

Organism	IGPS	PRAI
Neurospora crassa (filamentous ascomycete)	0.42	0.24
Saccharomyces cerevisiae (ascomycetous yeast)	0.39	0.23
Phycomyces blakesleanus (zygomycete)	0.45	0.30
Phanerochaete chrysosporium (basidiomycete)	0.46	0.28

[a] Number of aa matches between pairs of aligned sequences corrected for gaps were computed as described in the legend to Fig. 3 and divided by the length of the *Phytophthora* domain (280 aa for IGPS and 265 aa for PRAI).

Since the global structure of *TRP1* is so different from that encountered in other fungi, we examined the similarity between eukaryotic PRAI and IGPS enzymes at the aa sequence level (Table I). The results clearly demonstrate that PRAI domains diverged far more quickly than IGPS domains and they seem to show that both *Phytophthora* sequences are more similar to members of zygomycetes and to basidiomycetes than to ascomycetes. The difference in the evolutionary rate of IGPS in comparison to that of PRAI is remarkable since the enzymes catalyze consecutive reactions, possess similar structures and are often fused into two- and three-functional proteins. A similar situation was found in prokaryotes (Crawford, 1989).

To place *Phytophthora* IGPS and PRAI sequences within the context of fungal evolution, we constructed distance trees separately for each domain using all available complete eukaryotic sequences. The PRAI tree is shown in Fig. 3. The *Phytophthora* sequence is located outside the cluster which encompasses all fungi indicating that the lineage of oomycetes took a separate course of development before the speciation within the fungi began. The significance of this result is underlined by the fact that the IGPS tree exhibited an identical topology with that of PRAI (data not shown). These and other results (Gunderson et al., 1987) point out the need to reevaluate the phylogenetical relationship of oomycetes to fungi. The question whether IGPS and PRAI trees represent the real species tree can be answered only by analysing more sequences (Pamilo and Nei, 1988).

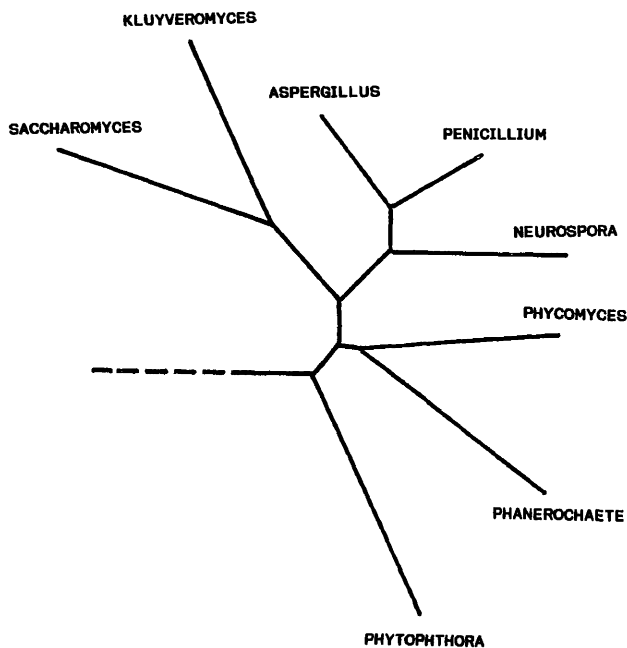

Fig. 3. Evolutionary distance tree for eukaryotic PRAI sequences. Sequence data were taken from Priestle et al. (1987), Choi et al. (1988), Schrank et al. (1991), Mullaney et al. (1985), Kos et al. (1988) and Stark and Milner (1989). PRAI and IGPS domains were localized by means of multiple alignment using the known three-dimensional structure of the *E. coli* enzyme (Priestle et al., 1987). Three protein segments located in sparsely occupied positions in the multiple alignment were deleted prior to analysis: a bridge of 13–15 aa between the C and F domains, the first 10 aa of the PRAI enzyme of *S. cerevisiae* (at the same position as the interdomain connection in multifunctional fungal enzymes) and 30 aa near the C-end of PRAI from *P. chrysosporium* (an intron-like sequence with no analogy in other PRAI sequences). To localize the root to the common ancestor, some prokaryotic and one archaebacterial sequence were included in the computation. Pair comparisons were made according to the method of Wilbur and Lipman (1983) using the program CLUSTAL1 (Higgins and Sharp, 1988) with the fixed gap penalty of three residues. The results were converted into distances and used as input in the program NJTREE (Studier and Keppler, 1988) based on the neighbor-joining method (Saitou and Nei, 1987) which was run on IBM3090.

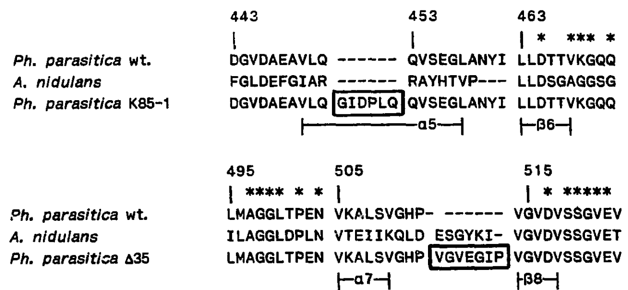

Fig. 4. Oligopeptide insertions tolerated in PRAI of *P. parasitica*. The aa are numbered from the start codon at nt 685 (see GenBank and EMBL accession No. M64473 for the *TRP1* nt sequence) and aligned with the sequence of *A. nidulans* (Mullaney et al., 1985; Kos et al., 1988). Asterisks indicate residues contributing to the active site in the *E. coli* enzyme or those fully conserved (Priestle et al., 1987). Horizontal lines show α-helices and β-strands deduced on the basis of sequence alignment with the *E. coli* enzyme and its x-ray structure (Priestle et al., 1987). Sequences introduced in vitro are boxed.

Surprisingly, the insertion Δ35 did not disturb *trpC*1117 complementation although it was located in the F domain (Fig. 1). Sequence analysis revealed that 21 extra bp (duplication of 11 bp and 10 bp remaining from Ω) coded for seven extra aa at this site. Another tolerated mutation within PRAI (K85-1) was constructed by cutting pTRΔ85 (Fig. 1) with *Bam*HI, filling in, and religating. This manipulation resulted in six extra aa (duplication 4 bp + insertion 10 bp + filling-in 4 bp). Fig. 4 shows the location of these oligopeptide insertions with respect to the secondary structure elements contributing to the eightfold βα barrel structure (Priestle et al., 1987). The insertion in pTRΔ35 is located at the N end of β-strand 8. Here a 5-aa deletion is evident in PRAI of *Ph. parasitica* as revealed by aligning with *A. nidulans*, *A. niger*, *P. chrysogenum* and *N. crassa* sequences (results not shown). This is similar to the situation found in bacteria and yeast (Priestle et al., 1987). Thus, the artificial Δ35 insertion imitates the situation normally found in filamentous fungi.

(d) Conclusions

(*1*) The *TRP1* gene of *P. parasitica* encodes the bifunctional protein PRAI–IGPS.

(*2*) The *IGPS* gene of *Phytophthora* is the only one of eight known eukaryotic *IGPS* genes which contains introns. In addition to the typical boundaries these introns possess the internal sequence CTAACT absent in most fungal introns (Gurr et al., 1987).

(*3*) Insertions of six and seven aa residues in PRAI were constructed that did not abolish its enzymatic activity.

(*4*) Phylogenetic distance trees for PRAI and IGPS sequences localized *Phytophthora* proteins outside the cluster comprising all other fungi for which these sequences are available.

ACKNOWLEDGEMENTS

We are obliged to Charles Yanofsky and Barbara J. Bachmann for providing us with bacterial strains, Frank Schubert (Max-Planck-Institut für Biophysikalische Chemie) for computer assistance with NJTREE, Desmond G. Higgins (Trinity College, Dublin) for making CLUSTAL programs available to us, Johanna Auer for excellent technical assistance and Margaret Nemec for improving the English. The experiments were partially completed at the Zentrales Isotopenlaboratorium für Biologische und Medizinische Forschung der Universität Göttingen. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

REFERENCES

Bucher, P. and Trifonov, E.N.: Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 24 (1986) 10009–10026.

Choi, H.T., Revuelta, J.L., Sadhu, C. and Jayaram, M.: Structural organization of the *TRP1* gene of *Phycomyces blakesleeanus*: implication for evolutionary gene fusion in fungi. Gene 71 (1988) 85–95.

Crawford, I.P.: Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. 43 (1989) 567–600.

Gunderson, J.H., Elwood, H., Ingold, A., Kindle, K. and Sogin, M.L.: Phylogenetical relationship between chlorophytes, chrysophytes, and oomycetes. Proc. Natl. Acad. Sci. USA 84 (1987) 5823–5827.

Gurr, S.J., Unkles, S.E. and Kinghorn, J.R.: The structure and organization of nuclear genes in filamentous fungi. In: Kinghorn J.R. (Ed.), Gene Structure in Eukaryotic Microbes. IRP Press, Oxford, 1987, pp. 93–139.

Higgins, D.G. and Sharp, P.M.: CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73 (1988) 237–244.

Hütter, R. and DeMoss, J.A.: Organization of the tryptophan pathway: a phylogenetic study of the fungi. J. Bacteriol. 94 (1967) 1896–1907.

Hütter, R. and Niederberger, P.: Tryptophan biosynthetic genes in eukaryotic microorganisms. Annu. Rev. Microbiol. 40 (1986) 55–77.

Karlovsky, P. and De Cock, A.W.A.M.: Buoyant density of DNA-Hoechst 33258 (Bisbenzimide) complexes in CsCl gradients: Hoechst 33258 binds to single AT base pairs. Anal. Biochem. 194 (1991) 192–197.

Kos, T., Kuijvenhoven, A., Hessing, H.G.M., Pouwels, P.H. and Van den Hondel, C.A.M.J.J.: Nucleotide sequence of the *Aspergillus niger trpC* gene: structural relationship with analogous genes of other organisms. Curr. Genet. 13 (1988) 137–144.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Mullaney, E.J., Hamer, J.E., Robert, K.A., Yelton, M.M. and Timberlake, W.E.: Primary structure of the *trpC* gene from *Aspergillus nidulans*. Mol. Gen. Genet. 199 (1985) 37–45.

Pamilo, P. and Nei, M.: Relationship between gene trees and species trees. Mol. Biol. Evol. 5 (1988) 568–583.

Prentki, P. and Krisch, H.M.: In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29 (1984) 303–313.

Priestle, J.P., Grütter, M.G., White, J.L., Vincent, M.G., Kania, M., Wilson, E., Jardetzky, T.S., Kirschner, K. and Jansonius, J.N.: Three-dimensional structure of the bifunctional enzyme *N*-(5′-phosphoribosyl)anthranilate isomerase–indole-3-glycerol-phosphate-synthase from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 84 (1987) 5690–5694.

Saitou, N. and Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 44 (1987) 406–425.

Schrank, A., Tempelaars, C., Sims, P.F.G., Oliver, S.G. and Broda, P.: The *trpC* gene of *Phanerochaete chrysosporium* is unique in containing an intron but nevertheless maintains the order of functional domains seen in other fungi. Mol. Microbiol. 5 (1991) 467–476.

Stark, M.J.R. and Milner, J.S.: Cloning and analysis of the *Kluyveromyces lactis TRP1* gene: a chromosomal locus flanked by genes encoding inorganic pyrophosphatase and histone H3. Yeast 5 (1989) 35–50.

Studier, J.A. and Kepler, K.J.: A note on the neighbor-joining method of Saitou and Nei. Mol. Biol. Evol. 5 (1988) 729–731.

Wilbur, W.J. and Lipman, D.J.: The context dependent comparison of biological sequences. SIAM J. Appl. Math. 44 (1984) 557–567.