RNA 3'-tail synthesis in *Streptomyces*: *in vitro* and *in vivo* activities of RNase PH, the SCO3896 gene product and polynucleotide phosphorylase

Patricia Bralley,¹ Bertolt Gust,^{2†} Samantha Chang,¹ Keith F. Chater² and George H. Jones¹

Correspondence

George H. Jones
george.h.jones@emory.edu

¹Department of Biology, Emory University, Atlanta, GA 30322, USA

²Department of Molecular Microbiology, The John Innes Centre, Norwich NR4 7UH, UK

As in other bacteria, 3'-tails are added post-transcriptionally to *Streptomyces coelicolor* RNA. These tails are heteropolymeric, and although there are several candidates, the enzyme responsible for their synthesis has not been definitively identified. This paper reports on three candidates for this role. First, it is confirmed that the product of *S. coelicolor* gene SCO3896, although it bears significant sequence similarity to *Escherichia coli* poly(A) polymerase I, is a tRNA nucleotidyltransferase, not a poly(A) polymerase. It is further shown that SCO2904 encodes an RNase PH homologue that possesses the polymerization and phosphorolysis activities expected for enzymes of that family. *S. coelicolor* RNase PH can add poly(A) tails to a model RNA transcript *in vitro*. However, disruption of the RNase PH gene has no effect on RNA 3'-tail length or composition in *S. coelicolor*; thus, RNase PH does not function as the RNA 3'-polyribonucleotide polymerase [poly(A) polymerase] in that organism. These results strongly suggest that the enzyme responsible for RNA 3'-tail synthesis in *S. coelicolor* and other streptomycetes is polynucleotide phosphorylase (PNPase). Moreover, this study shows that both PNPase and the product of SCO3896 are essential. It is possible that the dual functions of PNPase in the synthesis and degradation of RNA 3'-tails make it indispensable in *Streptomyces*.

Received 21 July 2005
Revised 14 October 2005
Accepted 1 November 2005

INTRODUCTION

Polyadenylation of the 3'-ends of RNAs, once thought to occur exclusively in eukaryotic cells, has now been shown definitively to occur in bacterial systems as well (Carpousis *et al.*, 1999; Rauhut & Klug, 1999; Sarkar, 1996, 1997). In *Escherichia coli*, polyadenylation targets RNAs for degradation (Carpousis *et al.*, 1999; Régnier & Arraiano, 2000). *E. coli* cells contain at least two enzymes capable of adding poly(A) tails to the 3'-ends of RNA molecules. The major polyadenylating enzyme, designated poly(A) polymerase I (PAP I, EC 2.7.7.19), was originally identified by virtue of its role in regulating plasmid copy number, and is a product of the *pcnB* gene (Cao & Sarkar, 1992; He *et al.*, 1993). Mutants of *E. coli* lacking PAP I still retain the ability to polyadenylate RNAs, indicating that there is at least one other polyadenylating enzyme in those cells (Kalapos *et al.*, 1994). That enzyme does not appear to be the product of the *f310* gene designated initially as PAP II, since overexpression of *f310* led neither to an increase in PAP activity in the relevant cells

nor to an increase in the level of RNA polyadenylation (Mohanty & Kushner, 1999a). Rather, the second PAP in *E. coli* is polynucleotide phosphorylase (PNPase, EC 2.7.7.8), which normally functions as a 3'-5'-exoribonuclease that catalyses the phosphorolysis of RNA chains, but under appropriate conditions *in vivo*, can degrade RNAs or synthesize poly(A) tails. PNPase is also responsible for the G, C and U residues that are found at low frequency in the poly(A) tails of RNAs from wild-type *E. coli* (Mohanty & Kushner, 2000b). The overall model for RNA degradation that has emerged from the studies in *E. coli* involves the generation of RNA 3'-ends via endonucleolytic cleavage by the endoribonucleases RNase E and RNase III, polyadenylation of the 3'-ends by PAP I (and under some conditions by PNPase) and degradation of the resulting RNAs exonucleolytically by PNPase (with the assistance of the DEAD box helicase RhlB), RNase II and RNase R (Cheng & Deutscher, 2005; Coburn & Mackie, 1999; Mohanty & Kushner, 1999b, 2000a; O'Hara *et al.*, 1995). This model applies especially to structured RNAs, with hairpins at their 3'-ends (Carpousis *et al.*, 1999; Régnier & Arraiano, 2000).

As a first step toward understanding the mechanisms of RNA degradation in streptomycetes, we demonstrated that a fraction of total RNA from *Streptomyces coelicolor* could be

[†]Present address: Pharmazeutische Biologie, Pharmazeutisches Institut, Universität Tübingen, 72076, Tübingen, Germany.

Abbreviations: PNPase, polynucleotide phosphorylase; PAP, poly(A) polymerase; TNT, tRNA nucleotidyltransferase.

bound by oligo(dT) cellulose and that this RNA fraction had the properties expected of RNA that was polyadenylated at its 3'-end (Bralley & Jones, 2001). Moreover, an enzyme activity capable of adding A residues to acceptor RNAs was demonstrated in *S. coelicolor* mycelial extracts. Unlike the situation in most eukaryotes, we observed that 16S and 23S (but not 5S) rRNAs as well as mRNAs were modified at their 3'-ends in *S. coelicolor* (Bralley & Jones, 2001). Most interestingly, when the nature of the 3'-tails associated with *S. coelicolor* RNAs was examined by cDNA cloning and DNA sequencing, it was observed that those tails were heteropolymeric, containing G, C and U as well as A residues (Bralley & Jones, 2002). Heteropolymeric tails were also observed at the 3'-ends of RNAs from *Streptomyces antibioticus* and we therefore believe that such tails are likely to be found in all streptomycetes (Bralley & Jones, 2003).

The presence of heteropolymeric tails at the 3'-ends of streptomycete RNAs raises the question of the nature of the enzyme responsible for synthesizing those tails. There are at least three candidates for that activity, based on the published sequence of the *S. coelicolor* genome (Bentley *et al.*, 2002). First, SCO3896 encodes a protein with significant sequence similarity to *E. coli* PAP I (Bralley & Jones, 2001). A second candidate PAP enzyme is RNase PH (EC 2.7.7.56: candidate gene SCO2904). Like PNPase, RNase PH is a 3'-5'-exoribonuclease that functions phosphorolytically (Deutscher *et al.*, 1988). RNase PH is involved in tRNA maturation in *E. coli* (Deutscher *et al.*, 1988), *Bacillus subtilis* (Wen *et al.*, 2005) and presumably in other bacteria as well. Because it is capable of polymerizing nucleoside diphosphates (Ost & Deutscher, 1990), it could conceivably

function as a PAP, in the same way that PNPase functions in *E. coli*. Finally, PNPase (candidate gene SCO5737) might catalyse the formation of RNA 3'-tails in *Streptomyces* as it does in *pcnB* mutants of *E. coli*.

In this report we confirm the results of Sohlberg *et al.* (2003) that the product of SCO3896 is not a PAP but is rather a tRNA nucleotidyltransferase (TNT), an enzyme that adds CCA to the ends of immature or damaged tRNAs. Moreover, we provide evidence that SCO3896 is an essential gene. We show further that SCO2904 encodes an RNase PH that like, polynucleotide phosphorylase, can polyadenylate RNA 3'-ends *in vitro*. However RNase PH appears to play no role in the synthesis or maintenance of poly(A) tails in *S. coelicolor* *in vivo*. We conclude, therefore, that the most likely candidate for the RNA 3'-polyribonucleotide polymerase in *S. coelicolor* is polynucleotide phosphorylase. PNPase is essential in *S. coelicolor*, presumably because of the critical role it plays in RNA degradation. The tRNA nucleotidyltransferase is also essential in *S. coelicolor*.

METHODS

Strains and media. *S. coelicolor* M145 and SCO2904 mutant strains were grown in Streptomyces minimal medium (SMM) as described by Kieser *et al.* (2000) with the exception that 1 % carboxymethyl cellulose (Sigma) replaced PEG 6000. *E. coli* strains were grown in LB medium supplemented with antibiotics as necessary.

Cloning of SCO3896 and SCO2904. SCO3896 and SCO2904 were cloned as PCR products using the primers listed in Table 1. In each case the forward primer used to amplify the gene replaced the GTG start codon annotated in the *S. coelicolor* genome database with ATG and the reverse primer introduced a *Bam*HI site downstream

Table 1. Primers used for PCR in this study

The *Nde*I and *Bam*HI sites are underlined in the SCTNT and SCRPH forward and reverse primers, respectively, used to amplify *S. coelicolor* SCO3896 and SCO2904. The bases underlined for the E19A.04, H24.18 and 3C3.23 forward and reverse primers correspond to sequences present in the disruption cassette used for PCR-targeted mutagenesis. The remaining bases represent flanking sequences for λ Red-mediated recombination.

Primer	Sequence
SCTNTFIN	5'-GTACCCTGTTTTCATATGCCGAACGCC-3'
SCTNTR1	5'-CGGAGGGTCGACGGATCCACCGGCTTCTAC-3'
SCRPHF1	5'-GCTAGGCTCACCCATATGTCAGGAATC-3'
SCRPHR1	5'-GCCTGCCGTTGGGGATCCCTTTACTCGCTT-3'
E19A.04.forw	5'-CGGCCCAGTTGCTGCCGTTGGCGTTCCTTTACTCGCTATTCCGGGGATCCGTCGACC-3'
E19A.04.rev	5'-GGACGGGCCCCCTTCCCCATGGGCGCTAGGCTCACCCGCTGTAGGCTGGAGCTGCTTC-3'
H24.18.forw	5'-GGGCGCCCGTCTCCGCCAAGGCCACGTACCCTGTTTTGTTATTCCGGGGATCCGTCGACC-3'
H24.18.rev	5'-TCATGCGTTGTCCGGGTGCTGAGCCCACCATTCCTTCAGTTGTAGGCTGGAGCTGCTTC-3'
3C3.23.forw	5'-GGCGTGGGACAAAAGACGACAAGCATCGGAGAAACCACTAATTCCGGGGATCCGTCGACC-3'
3C3.23.rev	5'-GTGGCGCTACGGGACGTCATCTGTCGGGGTCGTCCTTCTTGTAGGCTGGAGCTGCTTC-3'
AD20	5'-GGATCCGAATTCTCTAGAGC-3'
ADoT	5'-GGATCCGAATTCTCTAGAGCT ₁₇ -3'
16S1	5'-CTGTCGTCGCTCGTGTCTGTG-3'
PNP2	5'-GATCCAGGAGGACACCGGCGCCG-3'
PNP3	5'-CAGGCCTCCAGCCGCTCGCGTAG-3'

of the translation stop codon. The templates used were *S. coelicolor* cosmids H24 (*SCO3896* is orf 18 on the cosmid) and E19A (*SCO2904* is orf 04c on the cosmid). Conditions for PCR employed a 5 min hot start at 95 °C followed by 35 cycles consisting of denaturation at 95 °C for 1 min, annealing for 2 min at 55 °C and extension for 1 min at 72 °C. The last cycle was followed by a 10 min extension at 72 °C. The PCR products were purified by gel electrophoresis and the bands corresponding to *SCO3896* and *SCO2904* were isolated from the gel and digested with *Nde*I and *Bam*HI. The *SCO3896* fragment was ligated to the overexpression plasmid pET19B (Novagen) and the *SCO2904* fragment was ligated to pET11A; in each case the plasmid had been digested with *Nde*I and *Bam*HI. Ligation mixtures were used to transform *E. coli* DH5 α and transformants were screened by analysis of plasmid minipreps. The plasmids obtained by this analysis were designated pJSE3550 (*SCO3896*) and pJSE450 (*SCO2904*). A sample of each plasmid was used to transform the *E. coli* overexpression strain BL21(DE3)pLysS (Novagen).

Expression of *SCO3896* and *SCO2904* and protein purification. Overnight cultures (10 ml) of *E. coli* BL21(DE3)pLysS containing pJSE3550 or pJSE450 were used to inoculate two 500 ml cultures of LB medium containing carbenicillin (50 μ g ml⁻¹) and chloramphenicol (34 μ g ml⁻¹). The resulting cultures were grown to an OD₆₀₀ of 0.7 and IPTG was added to a final concentration of 1 mM. Cells were harvested from the 500 ml cultures 4 h after IPTG addition, washed with buffer I, composed of 50 mM Tris/HCl, pH 8.0, 5 mM MgCl₂, 5% (v/v) glycerol, 1 mM dithiothreitol, and stored frozen. For the purification of *SCO3896*, the cells were disrupted following the protocol that accompanies the Talon column procedure (BD Biosciences Clontech) and the protein was purified by Talon column chromatography following the manufacturer's instructions. Washed cells from the cultures bearing pJSE450 were disrupted as described previously (Jones *et al.*, 2003) and the resulting extract was applied to a 0.9 $\times$ 15 cm column of Source Q (Amersham Biosciences). The column was eluted at room temperature with a gradient of 0–0.3 M NaCl in buffer I and fractions containing the overexpressed protein were identified by SDS-PAGE of 20 μ l portions of every third fraction. Relevant fractions were pooled, brought to 1.5 M NaCl by the addition of 5 M NaCl and applied to a 1.2 $\times$ 12 cm column of Phenyl Sepharose CL-4B (Amersham Biosciences) equilibrated with buffer I containing 1.5 M NaCl. The column was washed successively at 4 °C with 50 ml buffer I/1.5 M NaCl, 50 ml buffer I/0.5 M NaCl and 100 ml buffer I. The enzyme began to elute in the 0.5 M NaCl wash and fractions containing *SCO2904* were pooled and concentrated by dialysis against ammonium sulfate without stirring (Schreier *et al.*, 1977). PNPase from *S. coelicolor* was prepared as described previously (Jones *et al.*, 2003).

Assays for PNPase, RNase PH, TNT and PAP. PNPase polymerization and phosphorolysis assays were performed essentially as described previously (Jones *et al.*, 2003) except that reaction mixtures contained 50 mM KCl. Phosphorolysis assays contained 1.97 mM total (³H-labelled plus unlabelled) poly(A). RNase PH assays were performed using the same conditions as those for PNPase. In both assays, mixtures were incubated for 15 min at 37 °C and [³H]ADP incorporation or the loss of acid-insoluble radioactivity was monitored by TCA precipitation and liquid scintillation counting. In some cases assay mixtures contained 150 μ g *E. coli* tRNA. Results are expressed as μ mol ADP incorporated or released based on the total amount of ADP (radioactive and non-radioactive) present in reaction mixtures.

TNT assays (60 μ l) contained 50 mM Tris/HCl, pH 8.5; 10 mM MgCl₂; 0.2 mM nonradioactive ATP or CTP; 20 μ Ci (740 kBq) ml⁻¹ of α -[³²P]ATP or CTP (3000 Ci mmol⁻¹, 110 TBq mmol⁻¹; Amersham Biosciences); 100 μ g bovine serum albumin ml⁻¹; 2 mg

E. coli tRNA (Sigma) ml⁻¹. Reaction mixtures were incubated for 15 min at 37 °C, after which 50 μ l samples were removed and the reaction products were collected by precipitation with trichloroacetic acid. Precipitates on glass fibre filters were analysed by liquid scintillation counting. *E. coli* PAP I was obtained from Ambion and assays were performed according to the supplier's instructions using tRNA as acceptor. In some experiments, reaction products from the TNT and PAP reactions were isolated by phenol extraction and ethanol precipitation and analysed on 10% denaturing polyacrylamide gels.

In other experiments, PNPase, RNase PH and PAP I were incubated with an RNA fragment derived from the readthrough transcript of the *rpsO–pnp* operon of *S. coelicolor*. This 333 base fragment was produced by RNase III cleavage of a transcript obtained from a cloned DNA fragment representing the *rpsO–pnp* intergenic region (Bralley & Jones, 2004). The transcript was synthesized using T7 RNA polymerase and α -[³²P]CTP, and the 3'-end of the RNase III digestion product represents the primary RNase III processing site in the *rpsO–pnp* intergenic region. Details of the processing of the *rpsO–pnp* transcript and the description of the fragments produced have been published elsewhere (Chang *et al.*, 2005). For the studies presented here, polymerization reaction mixtures were prepared as described above for PNPase, RNase PH and PAP I, except that tRNA was omitted. Instead, those mixtures contained 15 000 c.p.m. of the 333 base RNA fragment. Labelled RNAs were isolated from reaction mixtures by phenol extraction and ethanol precipitation and were examined by electrophoresis on 5% denaturing polyacrylamide gels along with Ambion Century-Plus size markers that were end-labelled using [³²P]pCp and T4 RNA ligase.

Analysis of oligo(A) stretches and RNA 3'-tail length and composition. Total RNA was prepared from *S. coelicolor* mycelium as described previously (Hsieh & Jones, 1995) and labelled at the 3'-end with [³²P]pCp and T4 RNA ligase (Bralley & Jones, 2001, 2003). Labelled RNAs were digested with RNases A and T1 and the digestion products were displayed on sequencing gels. This combination of RNases cleaves all phosphodiester bonds except those between adjacent A residues. End-labelled oligo(dT)₁₈ was used as a size standard in these experiments.

To determine RNA 3'-tail composition RT-PCR was carried out using 10 μ g total RNA as described previously (Bralley & Jones, 2002). Briefly, the oligo(dT)₁₇ primer, ADoT, listed in Table 1, was used for reverse transcription. A single round of PCR was utilized to amplify cDNA of 16S rRNA using the gene-specific forward primer 16S1 and the reverse primer AD20 (Table 1). Two rounds of semi-nested PCR were conducted using the gene-specific primers PNP2 and PNP3 (Table 1) in conjunction with primer ADoT. RT-PCR products were cloned using the TOPO-TA Cloning System (Invitrogen) and candidate clones were sequenced by the Emory University Sequencing Facility.

Gene disruptions. *SCO2904*, *SCO3896* and *SCO5737* were deleted by gene replacement using the PCR targeting method developed by Gust *et al.* (2003). A disruption cassette containing a resistance marker and an *oriT* (RK2) for efficient intergeneric conjugation was amplified by PCR with primers E19A.04.forw and E19A.04.rev (disruption of *SCO2904*), H24.18.forw and H24.18.rev (disruption of *SCO3896*), and 3C3.23.forw and 3C3.23.rev (disruption of *SCO5737*, Table 1) respectively. The PCR products were integrated into the corresponding cosmids by λ Red-mediated recombination. The cosmids were conjugated from *E. coli* ET12567/pUZ8002 to *S. coelicolor*, where they readily undergo homologous recombination.

To create a strain containing a second copy of *SCO3896*, a 2281 bp *Eco*RI–*Stu*I fragment of cosmid H24, bearing *SCO3896*, was cloned in pSET Ω . In pSET Ω the apramycin-resistance gene of pSET152 (Bierman *et al.*, 1992) is replaced with the spectinomycin/streptomycin-resistance marker (O'Connor *et al.*, 2002). The resulting construct

Fig. 1. SDS-PAGE of extracts from *E. coli* cultures overexpressing SCO3896 and SCO2904 and of the purified proteins. Lane 1, IPTG-induced culture of the strain bearing pJSE3550; lane 2, purified SCO3896, 5 µg; lane 3, IPTG-induced culture of the strain bearing pJSE450; lane 4, purified SCO2904, 5 µg; lane 5, molecular mass standards: phosphorylase *b* (105 kDa), bovine serum albumin (82 kDa), ovalbumin (49 kDa); carbonic anhydrase (33.3 kDa); soybean trypsin inhibitor (28.6 kDa); lysozyme (19.4 kDa). The gel was stained with Coomassie brilliant blue.

was transferred to *S. coelicolor* M145 by conjugation from *E. coli* and its integration was verified by PCR and Southern blotting. The resulting strain was then subjected to gene disruption using the REDIRECT

protocol (Gust *et al.*, 2003). To add a second copy of *pnp* to *S. coelicolor* M145, an approximately 3.1 kb PCR fragment, bearing *rpsO-pnp*, was prepared using cosmid 3C3 as a template. The PCR product was cloned into pSET152 as an *Xba*I fragment; that fragment was inserted into pSETΩ and the resulting construct was transferred to *S. coelicolor* M145 by conjugation.

Miscellaneous methods. Protein was determined by the method of Bradford (1976) using the Bio-Rad assay reagent with bovine serum albumin as a standard. SDS-PAGE was performed essentially as described by Laemmli (1970) and gels were stained with Coomassie brilliant blue.

RESULTS

SCO2904 encodes an RNase PH and SCO3896 a TNT

SCO2904 is annotated as an RNase PH and is 65 % identical, 71 % similar, over a stretch of 220 amino acids, to RNase PH from *E. coli*. We cloned SCO2904 in pET11A, overexpressed and purified the gene product and examined the RNase PH activity of the purified protein. As shown in Fig. 1, lane 3, IPTG induction led to the synthesis of a protein of the expected size (26 400 Da) and essentially homogeneous protein was obtained by the purification procedure described in Methods (Fig. 1, lane 4). Fig. 2 shows the results of polymerization and phosphorolysis assays of the SCO2904 gene product as compared with *S. coelicolor* PNPase. It is apparent that SCO2904 catalyses both the polymerization of ADP into an acid-insoluble product and the phosphorolysis of poly(A). Added tRNA had no effect on the polymerization activities of either PNPase or SCO2904. Based on its size, sequence similarity to the *E. coli* enzyme and enzymic activities, we conclude that SCO2904 encodes an RNase PH. The data of Fig. 2 indicate that the *S. coelicolor* RNase PH is less active than PNPase on a molar basis in both the polymerization and phosphorolysis reactions. This observation

Fig. 2. Polymerization and phosphorolysis assays of PNPase (●) and RNase PH (■). Polymerization assays were performed using [3 H]ADP as the substrate and phosphorolysis assays using [3 H]poly(A) as the substrate. Values are the mean of duplicate assays that differed by no more than 10 %, after subtracting a zero-time control from each experimental data point. In some experiments reaction mixtures contained 150 µg *E. coli* tRNA ml $^{-1}$.

may reflect the fact that RNase PH is thought to play a specific role in tRNA maturation in bacteria as opposed to the more general function of PNPase as a 3'-5'-exoribonuclease (Deutscher, 1995). Thus, RNase PH might not be expected to function as effectively as PNPase in phosphorylation or polymerization with poly(A) and ADP, respectively, as substrates.

SCO3896 is 36 % identical, 48 % similar to PAP I, the *E. coli* *pcnB* gene product (Bralley & Jones, 2001). PAPs and TNTs, the enzymes that add the CCA to the 3'-termini of immature and damaged tRNAs, belong to the same superfamily of nucleotidyltransferases (Holm & Sander, 1995; Yue *et al.*, 1996), so the SCO3896 gene product is also similar in sequence to the *E. coli* TNT (38 % identity, 48 % similarity), and it is even more similar to PAPS from *B. subtilis* (43 % identity, 53 % similarity). It was shown some years ago that PAPS is not the *B. subtilis* poly(A) polymerase; rather it is a tRNA nucleotidyltransferase (Raynal *et al.*, 1998). Sohlberg *et al.* (2003) have also shown that SCO3896 is a TNT and we have reached the same conclusion by overexpressing and purifying the gene product (Fig. 1, lanes 1 and 2) and verifying its activity. Fig. 3A shows that SCO3896 was

capable of adding C and A residues to tRNA acceptors (lanes 2 and 3) but did not synthesize the high-molecular-mass species that would be produced by a poly(A) polymerase (lane 4). It should be noted that the substrates for A and C addition in the total RNA pool used as the substrate in these assays are presumably species lacking complete CCA ends.

***S. coelicolor* RNase PH can polyadenylate RNA 3'-ends**

Given that *S. coelicolor* RNase PH has activities similar to those of PNPase, it was of interest to determine whether RNase PH, like PNPase, can add A residues to the 3'-ends of RNA molecules. To investigate this, we prepared an RNA fragment from a naturally occurring transcript, the read-through transcript of the *rpsO-pnp* operon of *S. coelicolor*. We synthesized a ^{32}P -labelled transcript representing the *rpsO-pnp* intergenic region, digested that transcript with RNase III and isolated a discrete digestion product 333 bases in length from the reaction. That product was then incubated in reaction mixtures prepared as described in Methods (but lacking tRNA) with PNPase, RNase PH or PAP I. The products of these reactions were isolated and analysed by

Fig. 3. Polyacrylamide gel analysis of reaction products. A. Autoradiogram of a 10 % denaturing polyacrylamide gel of the products of reactions of SCO3896 using *E. coli* tRNA as acceptor and α - ^{32}P ATP (lane 2) and $^3\alpha$ - ^{32}P CTP (lane 3) as donors. Lane 1 contains end-labelled RNA size standards. Lane 4 contains the products of a PAP I reaction with tRNA as the acceptor. B. Autoradiogram of a 5 % gel showing the products of reactions using the 333 base RNA fragment as an acceptor for polyadenylation. Lane 1 contains end-labelled size standards, lane 2 contains the ^{32}P -labelled 333 base fragment only (arrow), lanes 3–5 show the products of reaction of the 333 base fragment with PNPase, RNase PH and PAP I, respectively. The PNPase and RNase PH reactions contained unlabelled ADP and the PAP I reaction, unlabelled ATP. Reactions were performed and processed as described in Methods. Note that the 333 base fragment migrates anomalously. We believe that fragment possesses some secondary structure even in the presence of 7 M urea.

PAGE. Results of such analyses are shown in Fig. 3B. As expected, PAP I, in the presence of ATP, converted the 333 base fragment to larger species, some of which were over 1000 bases in length (Fig. 3B, lane 5). Similarly, PNPase, when incubated with the 333 base fragment and ADP, converted the fragment into higher-molecular-mass products (Fig. 3B, lane 3). Fig. 3B, lane 4, shows that RNase PH also has this polymerizing activity. The 333 base substrate was converted to species at least 750 bases in total length. Thus, RNase PH, like PNPase, can function as a poly(A) polymerase *in vitro*. To our knowledge this is the first demonstration that RNase PH has poly(A) polymerase activity with a naturally occurring substrate other than tRNA.

RNase PH is not the *S. coelicolor* PAP

Given the observation that *S. coelicolor* RNase PH can function as a PAP *in vitro* it was of considerable interest to examine the potential role of this enzyme in the synthesis of RNA 3'-tails *in vivo*. To this end, *SCO2904* was disrupted by the insertion of a viomycin-resistance cassette using the REDIRECT gene disruption protocol (Gust *et al.*, 2003). Disruption of the gene was confirmed by PCR and by Southern blotting (data not shown), demonstrating insertion of the viomycin-resistance cassette into *SCO2904*. RNA was prepared from *S. coelicolor* M145 and from two putative disruptant strains, JSE450A and 450B, and the 3'-tails associated with the RNA species in those preparations were analysed for oligo(A) stretches embedded in the heteropolymeric 3'-tails (Fig. 4). It can be seen that the distribution of those stretches is identical in the parent strain, *S. coelicolor* M145, and in the two disruptant strains. Prominent bands representing oligo(A) stretches of 1–18 residues are observed in the preparations from the parent strain and both disruptant strains. Overexposure of the autoradiogram shown in Fig. 4 revealed bands in all three preparations that were 25–35 residues in length. This observation confirms previous results that while oligo(A) stretches of less than 18–20 residues are most commonly observed as components of the heteropolymeric 3'-tails of streptomycete RNAs, longer stretches do occur at a lower frequency (Bralley & Jones, 2003). It should be noted that because the poly(A) tails of streptomycete RNAs are highly heteropolymeric (Bralley & Jones, 2002), the pattern of continuous increase in tail length observed in *E. coli* (lane 3) is not observed for the tails from *S. coelicolor*.

To assess further the effect of RNase PH upon 3'-tail formation, cDNA clones of 16S rRNA and several mRNAs were generated by oligo(dT)-dependent RT-PCR using RNA from the mutant JSE450A. The sequences of these clones, shown in Table 2, revealed no significant difference in tail length or composition of cDNAs corresponding to the polyadenylated RNAs isolated from wild-type or mutant strains. The tails of 16S rRNA appeared similar to 16S rRNA tails isolated from *S. coelicolor* M145 (Bralley & Jones, 2002). These experiments generated a variety of cDNAs from genes possessing unanticipated homology to the primer (PNP3) used in RT-PCR. These cDNAs, from *leuA* (*SCO387*), *clpP*

Fig. 4. Measurement of the lengths of oligo(A) stretches found in the 3'-tails of RNAs isolated from *S. coelicolor* and *E. coli* strains. Total RNA was labelled and digested with RNases A and T1 and the digestion products were fractionated on a 12% polyacrylamide gel. Lanes 1 and 2, two dilutions of end-labelled oligo(dT)₁₈; lane 3, RNA from *E. coli* DH5 α ; lane 4, RNA from *S. coelicolor* M145; lanes 5 and 6, RNAs from JSE450A and JSE450B, respectively; lane 7, control reaction in which RNA was omitted.

(*SCO3044*) and a 23S rRNA, were isolated from both wild-type and mutant strains. These clones revealed tails with compositions similar to those previously isolated from genes *actII-orf4* and *redD* (Bralley & Jones, 2002) and *S. antibioticus pnp* (Bralley & Jones, 2003). The fact that both RNA tail length and composition in the mutant JSE450A appear similar to that found in M145 strongly suggests that RNase PH is not involved in the synthesis of RNA 3'-tails in *S. coelicolor*.

TNT and PNPase are essential proteins in *S. coelicolor*

As has been reported previously, numerous attempts by us and by others to disrupt *pnp* in *S. coelicolor* and in *S.*

Table 2. Attachment sites and composition of heteropolymeric poly(A) tail sequences in *S. coelicolor* M145 and *S. coelicolor* M145 *rph*⁻ (JSE450A)

Clone	Gene*	Attachment site*	cDNA tail sequence
M145 clones[†]			
GHJ 412	16S rRNA (1528)	1382	TCA ₂ GA ₂ GA ₁₀ TA ₈
GHJ 405		1379	TGATA ₇ TAT ₂ ACTATA ₂ CATA ₈ TA ₁₇
GHJ 411		1377	GCA ₁₉
GHJ 1		1528	A ₁₂
GHJ 2		1386	AGTGAGA ₂ GA ₅ GA ₂ GA ₅ N ₂ CA ₂ GA ₁₅ TA ₃
GHJ 3	<i>leuA</i> (2350)	1374	A ₁₅
GHJ 419		1521	A ₁₇
GHJ 420		616	A ₂ GA ₂₁
GHJ 5		933	CATACT ₂ A ₅ TA ₁₆
GHJ 6		899	A ₂ TA ₁₉
GHJ 7	<i>redD</i> (1053)	870	AT ₂ GATGA ₂₃
GHJ 8		784	A ₂ TA ₁₇
GHJ 9		692	TA ₂ CA ₂ TA ₁₈
M145 <i>rph</i>⁻ clones			
GHJ 416	16S rRNA (1528)	1399	TA ₁₇
GHJ 404		1384	A ₂ CA ₄ GATA ₁₇
GHJ 418		1381	A ₂ GATACTAT ₂ CATACA ₈ TAGA ₁₇
GHJ 417		1377	GTGTGACG ₃ A ₃ ACA ₂₂
GHJ 413		1376	GA ₂₂
GHJ 403	<i>clpP</i> (656)	1375	TAGA ₂ GCCA ₁₇
GHJ 421		587	A ₂ TGAGA ₄ G ₂ AGCA ₃ C ₂ AGA ₂ GAG ₂ AG ₃ - AG ₅ C ₂ AG ₂ TG ₂ A ₄ TA ₄ CA ₁₅
GHJ 401		352	GA ₃ CGA ₁₈ GA ₂
GHJ 422		291	TGCGA ₂ GA ₂₀
GHJ 423		2955	A ₂₆

*The position of the 3'-tail was reckoned from the mature 5'-end of the rRNA gene or from the translation start site of mRNAs; the numbers in parentheses refer to the last nucleotide of the mature 3'-end of 16S rRNA or the translational stop codon.

†Clones GHJ 1–GHJ 9 were described by Bralley & Jones (2002).

antibioticus have been unsuccessful (Bralley & Jones, 2001, 2002, 2003). Attempts were also made, as a part of the studies reported here, to disrupt *SCO3896* and *pnp* (*SCO5737*) using a PCR-targeted mutagenesis approach that yields high frequencies of gene disruption in *Streptomyces* (Gust *et al.*, 2003). To this end, an apramycin-resistance cassette was synthesized by PCR with 5'- and 3'-ends that correspond to regions flanking either *SCO3896* or *SCO5737*. These cassettes were inserted into the corresponding cosmids and the resulting *E. coli* strains bearing the mutagenized cosmids were mixed with spores of *S. coelicolor* M145 to allow conjugal transfer of the cosmids. Subsequent homologous recombination would be expected to result in replacement of the chromosomal copy of *SCO3896* or *SCO5737* with the disrupted copy at high frequencies. More than 1000 exconjugant colonies were screened from each conjugation by PCR, but they all contained both a disrupted and an intact copy of *SCO3896* and

SCO5737, respectively. No double crossover replacements were observed.

This result suggested that both *SCO3896* and *SCO5737* are essential in *S. coelicolor*. To test this possibility, a copy of each gene was cloned into the integrative plasmid vector pSETΩ and introduced into M145. pSETΩ contains the bacteriophage φC31 *attP* site and integrates into the *attB* site in the *S. coelicolor* chromosome (Bierman *et al.*, 1992). The insertions into the *S. coelicolor* genome were verified by Southern blotting and the attempts to disrupt *SCO3896* and *SCO5737* were repeated. In the case of *SCO3896*, more than 30 % of the exconjugants obtained were double crossovers and approximately 50 % double crossovers were observed in the case of *SCO5737*. Both PCR using primers specific for each gene, and Southern blotting, confirmed that the original chromosomal copy had been successfully disrupted in some of those exconjugants. Thus, it was possible to disrupt

SCO3896 or SCO5737 only after inserting a second copy of the relevant gene into the chromosome.

DISCUSSION

As is the case in *E. coli*, the polyadenylation of the 3'-ends of RNA molecules in *Streptomyces* is expected to play an important role in the degradation of those RNAs. *S. coelicolor* contains several of the enzymes known to be involved in degradation of RNAs in *E. coli*, including RNase E and RNase III (Hagège & Cohen, 1997; Price *et al.*, 1999). Unlike the situation in *E. coli*, streptomycetes do not appear to contain an RNase II, the 3'-5'-exoribonuclease that functions hydrolytically or RNase R (Zuo & Deutscher, 2001). Thus, the major 3'-5'-exoribonuclease in *S. coelicolor* is almost certainly PNPase. Lee & Cohen (2001) have provided data suggesting that in *S. coelicolor*, PNPase and RNase E may be organized into a supramolecular complex like the *E. coli* degradosome.

Several recent studies of the streptomycete PNPase have confirmed its function as a phosphorolytic exoribonuclease (Jones & Bibb, 1996; Jones *et al.*, 2003). The unanswered question has to do with its function in the synthesis of RNA 3'-tails. The most straightforward way to examine this question would be to disrupt *pnp* in *S. coelicolor* and to examine the effects of that disruption on the length and composition of the tails associated with RNAs in the disruptant strains. However, as we demonstrate here, PNPase is an essential gene in *S. coelicolor*. Thus, we have adopted a 'process of elimination' approach to define the role of PNPase in RNA 3'-end synthesis. We have shown that SCO3896, which has significant sequence similarity to *E. coli* PAP I, is not a poly(A) polymerase but is rather a tRNA nucleotidyltransferase. These findings confirm the results of Sohlberg *et al.* (2003).

As PNPase would function in its polymerization mode as a polyribonucleotide polymerase, it seemed formally possible that a second enzyme with similar activity, RNase PH, might be responsible for synthesis of RNA 3'-tails in streptomycetes. Indeed we have shown that *S. coelicolor* RNase PH can polymerize ADP *in vitro* just as can PNPase. Moreover, we have shown that RNase PH can add 3'-tails to a cellular RNA, again like PNPase. Thus, the analysis of Sohlberg *et al.* (2003), which was based on their observation that PNPase could function as a PAP *in vitro*, omitted another enzyme, RNase PH, which has exactly the same *in vitro* activity. We show here, however, that disruption of the RNase PH gene has no effect on the length or composition of the tails found at the 3'-ends of various cellular RNAs. Thus, we conclude that neither the product of SCO3896 nor that of SCO2904 plays a significant role in the synthesis of the 3'-tails in *S. coelicolor*. We argue, as do Sohlberg *et al.* (2003), that the enzyme responsible for the synthesis of those tails is PNPase. This conclusion is consistent with the observation that spinach chloroplasts and a *Synechocystis* species, like *S. coelicolor*, do not encode a dedicated PAP (Rott *et al.*, 2003;

Yehudai-Resheff *et al.*, 2001). In the two former systems, and in *pcnB* mutants of *E. coli*, the RNA 3'-polyribonucleotide polymerase is PNPase and as in *S. coelicolor*, the 3'-tails of RNAs in spinach chloroplasts, in the *Synechocystis* sp. and in *pcnB* mutants are heteropolymeric (Mohanty & Kushner, 2000b; Rott *et al.*, 2003; Yehudai-Resheff *et al.*, 2001). We cannot exclude the possibility that RNase PH may add residues infrequently to RNA 3'-tails *in vivo*, as is the case for PNPase (Mohanty & Kushner, 2000b), but our data clearly indicate that RNase PH is not the PAP in *S. coelicolor*.

Why should PNPase and TNT be essential in *S. coelicolor*? In the case of PNPase, the answer would appear to be that it is the only enzyme in streptomycetes capable of catalysing two important reactions involved in RNA degradation: the exonucleolytic removal of 3'-tails (recall that streptomycetes do not contain an RNase II or RNase R) and the synthesis of those same tails. As regards SCO3896, inspection of the *S. coelicolor* genome reveals that less than a third of the tRNA genes encode the CCA end. Thus, most *S. coelicolor* tRNAs, many of which must be essential for protein synthesis, require SCO3896 to produce mature 3'-ends. In contrast, the studies of Zhu & Deutscher (1987) show definitively that TNT is not essential in *E. coli*. Thus, all essential tRNA genes in that organism must encode the CCA end.

We show in another publication that of the known bacterial phyla only the β - and γ -proteobacteria contain both a TNT and a dedicated PAP (Bralley *et al.*, 2005). We believe it possible, therefore, that PNPase may function as the RNA 3'-polyribonucleotide polymerase in other bacterial species. Our recent results suggest, however, that there are at least three biochemical systems for the synthesis of RNA 3'-tails in bacteria. As discussed above, the PAP I homologue in *B. subtilis* is actually a TNT. Moreover, we have demonstrated poly(A) tail synthesis in *B. subtilis* mutants lacking both the TNT and PNPase (Campos-Guillén *et al.*, 2005). We are currently examining the enzymology of poly(A) tail formation in *B. subtilis*.

ACKNOWLEDGEMENTS

This work was supported by grant no. MCB-0133520 from the National Science Foundation to G. H. J. and by 208/IGF12432 from the Biotechnology and Biological Research Council to K. F. C.

REFERENCES

- Bentley, S. D., Chater, K. F., Cerdano-Tarraga, A. M. & 40 other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.
- Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N. & Shoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.

- Bralley, P. & Jones, G. H. (2001). Poly(A) polymerase activity and RNA polyadenylation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**, 1155–1164.
- Bralley, P. & Jones, G. H. (2002). cDNA cloning confirms the polyadenylation of RNA decay intermediates in *Streptomyces coelicolor*. *Microbiology* **148**, 1421–1425.
- Bralley, P. & Jones, G. H. (2003). Overexpression of the polynucleotide phosphorylase gene (*pnp*) of *Streptomyces antibioticus* affects mRNA stability and poly(A) tail length but not ppGpp levels. *Microbiology* **149**, 2173–2182.
- Bralley, P. & Jones, G. H. (2004). Organization and expression of the polynucleotide phosphorylase gene (*pnp*) of *Streptomyces*: processing of *pnp* transcripts in *Streptomyces antibioticus*. *J Bacteriol* **186**, 3160–3172.
- Bralley, P., Chang, S. A. & Jones, G. H. (2005). A phylogeny of bacterial nucleotidyltransferases: *Bacillus halodurans* contains two tRNA nucleotidyltransferases. *J Bacteriol* **187**, 5927–5936.
- Campos-Guillén, J., Bralley, P., Jones, G. H., Bechhofer, D. H. & Almeda-Alvarez, G. (2005). Addition of poly(A) and heteropolymeric 3'-ends in *Bacillus subtilis* wild-type and PNPase-deficient strains. *J Bacteriol* **187**, 4698–4706.
- Cao, G. J. & Sarkar, N. (1992). Identification of the gene for an *Escherichia coli* poly(A) polymerase. *Proc Natl Acad Sci U S A* **89**, 10380–10384.
- Carpousis, A. J., Vanzo, N. F. & Raynal, L. C. (1999). mRNA degradation: a tale of poly(A) and multiprotein machines. *Trends Genet* **15**, 24–28.
- Chang, S. A., Bralley, P. & Jones, G. H. (2005). The *absB* gene encodes a double-strand specific endoribonuclease that cleaves the readthrough transcript of the *rspO-pnp* operon in *Streptomyces coelicolor*. *J Biol Chem* **280**, 33213–33219.
- Cheng, Z. F. & Deutscher, M. P. (2005). An important role for RNase R in mRNA decay. *Mol Cell* **17**, 313–318.
- Coburn, G. A. & Mackie, G. A. (1999). Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. *Prog Nucleic Acids Res Mol Biol* **62**, 55–105.
- Deutscher, M. (1995). tRNA processing nucleases. In *tRNA: Structure, Biosynthesis and Function*, pp. 51–65. Edited by D. Söll & U. L. RajBhandary. Washington, DC: American Society for Microbiology.
- Deutscher, M., Marshall, G. T. & Cudny, H. (1988). RNase PH: an *Escherichia coli* phosphate-dependent nuclease distinct from polynucleotide phosphorylase. *Proc Natl Acad Sci U S A* **85**, 4710–4714.
- Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**, 1541–1546.
- Hagège, J. M. & Cohen, S. N. (1997). A developmentally regulated *Streptomyces* endoribonuclease resembles ribonuclease of *Escherichia coli*. *Mol Microbiol* **25**, 1077–1090.
- He, L., Soderbom, F., Wagner, E. G., Binnie, U., Binns, N. & Masters, M. (1993). PcnB is required for the rapid degradation of RNAI, the antisense RNA that controls the copy number of ColE1-related plasmids. *Mol Microbiol* **9**, 1131–1142.
- Holm, L. & Sander, C. (1995). DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem Sci* **20**, 345–347.
- Hsieh, C.-J. & Jones, G. H. (1995). Nucleotide sequence, transcriptional analysis and glucose regulation of the phenoxazinone synthase gene from *Streptomyces antibioticus*. *J Bacteriol* **177**, 5740–5747.
- Jones, G. H. & Bibb, M. J. (1996). Guanosine pentaphosphate synthetase from *Streptomyces antibioticus* is also a polynucleotide phosphorylase. *J Bacteriol* **178**, 4281–4288.
- Jones, G. H., Symmons, M. F., Hankins, J. S. & Mackie, G. A. (2003). Overexpression and purification of untagged polynucleotide phosphorylases. *Protein Expr Purif* **32**, 202–209.
- Kalapos, M. P., Cao, G. J., Kushner, S. R. & Sarkar, N. (1994). Identification of a second poly(A) polymerase in *Escherichia coli*. *Biochem Biophys Res Commun* **198**, 459–465.
- Kieser, Y., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. (2000). *Practical Streptomyces Genetics*. Norwich: John Innes Foundation.
- Laemmli, U. (1970). Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lee, K. & Cohen, S. N. (2001). A *Streptomyces coelicolor* functional analog of *E. coli* RNase E shows evolutionary shuffling of catalytic and PNPase domains. In *Proceedings of the 12th International Symposium on the Biology of Actinomycetes*, p. 66.
- Mohanty, B. K. & Kushner, S. R. (1999a). Residual polyadenylation in poly(A) polymerase I (*pcnB*) mutants of *Escherichia coli* does not result from the activity encoded by the *f310* gene. *Mol Microbiol* **34**, 1109–1119.
- Mohanty, B. K. & Kushner, S. R. (1999b). Analysis of the function of *Escherichia coli* poly(A) polymerase in RNA metabolism. *Mol Microbiol* **34**, 1094–1108.
- Mohanty, B. K. & Kushner, S. R. (2000a). Polynucleotide phosphorylase, RNase II and RNase E play different roles in the *in vivo* modulation of polyadenylation in *Escherichia coli*. *Mol Microbiol* **36**, 982–994.
- Mohanty, B. K. & Kushner, S. R. (2000b). Polynucleotide phosphorylase functions both as a 3'-5' exonuclease and a poly(A) polymerase in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 11966–11971.
- O'Connor, T. J., Kanellis, P. & Nodwell, J. R. (2002). The *ramC* gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol Microbiol* **45**, 45–57.
- O'Hara, E. B., Chekanova, J. A., Ingle, C. A., Kushner, Z. R., Peters, E. & Kushner, S. R. (1995). Polyadenylation helps regulate mRNA decay in *Escherichia coli*. *Proc Natl Acad Sci U S A* **92**, 1807–1811.
- Ost, K. A. & Deutscher, M. (1990). RNase PH catalyzes a synthetic reaction, the addition of nucleotides to the 3'-ends of RNA. *Biochimie* **72**, 813–818.
- Price, B., Adamis, T. & Champness, W. (1999). A *Streptomyces coelicolor* antibiotic regulatory gene, *absB*, encodes an RNase III homolog. *J Bacteriol* **181**, 6142–6151.
- Rauhut, R. & Klug, G. (1999). mRNA degradation in bacteria. *FEMS Microbiol Rev* **23**, 353–370.
- Raynal, L. C., Krusch, H. M. & Carpousis, A. J. (1998). The *Bacillus subtilis* nucleotidyl transferase is a tRNA CCA-adding enzyme. *J Bacteriol* **180**, 6276–6282.
- Régnier, P. & Arraiano, C. M. (2000). Degradation of mRNA in bacteria: emergence of ubiquitous features. *Bioessays* **22**, 235–244.
- Rott, R., Zipor, G., Portnoy, V., Liveanu, V. & Schuster, G. (2003). RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from *E. coli*. *J Biol Chem* **278**, 15771–15777.
- Sarkar, N. (1996). Polyadenylation of mRNA in bacteria. *Microbiology* **142**, 3125–3133.
- Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. *Annu Rev Biochem* **66**, 173–197.
- Schreier, M. H., Erni, B. & Staehlin, T. (1977). Initiation of mammalian protein synthesis. I. Purification and characterization of seven initiation factors. *J Mol Biol* **116**, 727–753.
- Sohlberg, B., Huang, J. & Cohen, S. N. (2003). The *Streptomyces coelicolor* polynucleotide phosphorylase homologue, and not the

putative poly(A) polymerase can polyadenylate RNA. *J Bacteriol* **185**, 7273–7278.

Wen, T., Oussenko, I. A., Pellegrini, O., Bechhofer, D. H. & Condon, C. (2005). Ribonuclease PH plays a major role in the exonucleolytic maturation of CCA-containing tRNA precursors in *Bacillus subtilis*. *Nucleic Acids Res* **33**, 3636–3643.

Yehudai-Resheff, S., Hirsh, M. & Schuster, G. (2001). Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. *Mol Cell Biol* **21**, 5408–5416.

Yue, D., Maizels, N. & Weiner, A. M. (1996). CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the hyperthermophile *Sulfolobus shibatae*. *RNA* **2**, 895–908.

Zhu, L. & Deutscher, M. P. (1987). tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. *EMBO J* **6**, 2473–2477.

Zuo, Y. & Deutscher, M. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29**, 1017–1026.