

# Two different thymidylate kinase gene homologues, including one that has catalytic activity, are encoded in the onion yellows phytoplasma genome

Shin-ichi Miyata,<sup>1</sup> Kenro Oshima,<sup>1</sup> Shigeyuki Kakizawa,<sup>1</sup>  
Hisashi Nishigawa,<sup>2</sup> Hee-Young Jung,<sup>2</sup> Tsutomu Kuboyama,<sup>3</sup>  
Masashi Ugaki<sup>1,2</sup> and Shigetou Namba<sup>1,2</sup>

Correspondence  
Shigetou Namba  
snamba@ims.u-tokyo.ac.jp

<sup>1</sup>Division of Integrated Biosciences, The University of Tokyo, 202 Bioscience Bldg, 5-1-5 Kashiwanoha, Chiba 277-8562, Japan

<sup>2</sup>Division of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

<sup>3</sup>Department of Agricultural Sciences, School of Agriculture, Ibaraki University, Chuo 3-21-1, Ami-machi, Inashiki-gun, Ibaraki 300-0393, Japan

Received 13 March 2003  
Revised 28 April 2003  
Accepted 6 May 2003

Thymidylate kinase (TMK) catalyses the phosphorylation of dTMP to form dTDP in both the *de novo* and salvage pathways of dTTP synthesis in both prokaryotes and eukaryotes. Two homologues of bacterial thymidylate kinase genes were identified in a genomic library of the onion yellows (OY) phytoplasma, a plant pathogen that inhabits both plant phloem and the organs of insects. Southern blotting analysis suggested that the OY genome contained one copy of the *tmk-b* gene and multiple copies of the *tmk-a* gene. Sequencing of PCR products generated by amplification of *tmk-a* enabled identification of three other copies of *tmk-a*, although the ORF in each of these was interrupted by point mutations. The proteins, TMK-a and TMK-b, encoded by the two intact genes contained conserved motifs for catalytic activity. Both proteins were overexpressed as fusion proteins with a polyhistidine tag in *Escherichia coli* and purified, and TMK-b was shown to have thymidylate kinase activity. This is believed to be the first report of the catalytic activity of a phytoplasmal protein, and the OY phytoplasma is the first bacterial species to be found to have two intact homologues of *tmk* in its genome.

## INTRODUCTION

Phytoplasmas (formerly called mycoplasma-like organisms or MLOs) are members of the class *Mollicutes* that cause several hundred plant diseases (Doi *et al.*, 1967; Lee *et al.*, 2000). They are transmitted from plant to plant by insect vectors (mostly leafhoppers) and cause various symptoms, including stunting (shortened internodes and dwarfed leaves and flowers), yellowing, proliferation (witches' broom), virecence (greening of the floral tissue) and phyllody (leaf-like petals and sepals) (McCoy *et al.*, 1989; Kirkpatrick, 1992; Shiomi *et al.*, 1996). Currently antibiotic therapy with tetracycline is the only means of curing phytoplasma disease (Davies & Clark, 1994), and a better understanding of the cellular metabolism of phytoplasmas is essential for

developing novel anti-phytoplasma therapeutics. Unfortunately, the inability to culture phytoplasmas *in vitro* has inhibited the study of their biology (Lee *et al.*, 2000).

Thymidylate kinase (TMK, EC 2.7.4.9), which catalyses the transfer of a terminal phosphoryl group from ATP to dTMP, is crucial to both the *de novo* synthetic and the salvage pathways for pyrimidine deoxyribonucleotides. However, mollicutes have only the salvage pathway (Pollack *et al.*, 1997). The cascade of enzymes that convert nucleosides to nucleoside triphosphates and deoxynucleoside triphosphates is of interest in drug discovery, as these molecules are vital precursors in the synthesis of DNA, RNA, and other cellular macromolecules. TMK has been extensively studied as an antiviral target. The variability in the active-site residues and catalytic properties of TMKs in different organisms (viral, eukaryotic and bacterial) has opened the possibility of designing specific and selective inhibitors (Jong & Campbell, 1984; Darby, 1995; Griffiths, 1995; Lavie *et al.*, 1998). The necessity of TMK for bacterial growth (Li *et al.*, 2000), and the presence of the enzyme in both Gram-negative and

Abbreviations: OY, onion yellows; TMK, thymidylate kinase.

The GenBank accession numbers for the *tmk-a* and *tmk-b* genes are AB010446 and AB094668, respectively. The accession numbers of *tmk-a* homologues T01, T03 and T08 are AB100419, AB100420 and AB100421, respectively.

Gram-positive bacterial pathogens, make it an attractive target for the development of novel, broad-spectrum anti-bacterial agents.

TMK is largely encoded as a single-copy gene in bacteria for which the genomes have been completely sequenced [refer to the Clusters of Orthologous Groups of proteins (COGs) database: <http://www.ncbi.nlm.nih.gov/COG/index.html>], and *tmk* genes have been annotated in the genomes of *Mycoplasma genitalium* (Fraser *et al.*, 1995), *Mycoplasma pneumoniae* (Himmelreich *et al.*, 1996), *Mycoplasma pulmonis* (Chambaud *et al.*, 2001) and *Ureaplasma urealyticum* (Glass *et al.*, 2000).

TMK enzymes are globular dimeric proteins with a folding pattern similar to that of nucleoside monophosphate kinases (Ostermann *et al.*, 2000), which possess three loops crucial for enzyme activity: the phosphate-binding motif at the N-terminus (P-loop), the nucleoside-monophosphate-binding domain, and the region that covers part of the P-loop upon substrate binding (the LID domain).

With the aim of developing an effective therapy for phytoplasmosis, we focused on the catalytic activity of TMK, which may be one of the important and active cellular metabolic enzymes in phytoplasmas. Thus we tried to isolate the *tmk* gene from the onion yellows (OY) phytoplasma genome, and examined the genomic organization of *tmk* genes and the catalytic activity of phytoplasma TMK.

## METHODS

**Phytoplasma lines.** OY-W, the wild-type line of OY phytoplasma, was isolated in 1982 from an onion, *Allium cepa*, in Saga Prefecture, Japan (Shiomi *et al.*, 1996). It has since been maintained in a plant host, garland chrysanthemum, *Chrysanthemum coronarium*, using a leafhopper vector, *Macrostelus striifrons*. OY-W produces a wide variety of symptoms in its plant hosts, including virescence, yellowing, phyllody, stunting, proliferation, and witches' broom. OY-M, a mildly pathogenic line, was isolated after maintaining OY-W for 11 years (Shiomi *et al.*, 1998). Plants infected with OY-M do not have shortened internodes, and other symptoms of disease appear mild. The reproducibility and stability of the symptoms of OY-M infected plants have been confirmed more than 60 times over 4 years (Oshima *et al.*, 2001).

**Extraction of phytoplasma genomic DNA, construction of a genomic library, DNA sequencing, and homology analysis.** Phytoplasma-enriched DNA from infected plants was extracted using a previously reported procedure (Kuboyama *et al.*, 1998; Miyata *et al.*, 2002a). Purified OY genomic DNA was completely digested with the restriction endonuclease *Hind*III and ligated into pUC18. The ligated genomic DNA was used to transform *E. coli* JM109, and several thousand plasmid clones were obtained. The cloned DNA fragments were then excised from pUC18 and prepared as probes for dot-blot hybridization to DNA samples extracted from OY-infected and healthy plants (Kuboyama *et al.*, 1998). DNA probes were labelled with glutaraldehyde and the hybridization products were detected using the ECL direct nucleic acid labelling and detection system (Amersham Biosciences), following the manufacturer's protocols. Approximately 200 positive clones were obtained.

The DNA inserts were sequenced using the dideoxynucleotide chain

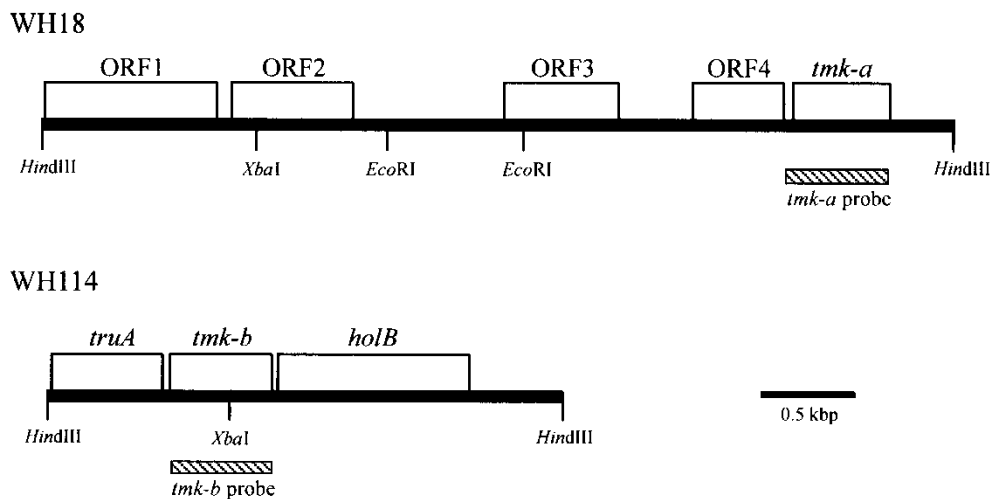
termination method with an automated DNA sequencer (model 377, Applied Biosystems). The similarity of each ORF to known genes was analysed using sequence interpretation tools (Institute for Chemical Research, University of Kyoto, Japan) and the BLAST algorithm (Altschul *et al.*, 1990) via the GenomeNet server (<http://www.genome.jp/>). The amino acid sequences of proteins that were similar to each ORF were aligned using CLUSTAL W version 1.7 (Thompson *et al.*, 1994). The sequences used for comparison were TMK of *M. genitalium*, AAC71222, *M. pneumoniae*, AE000016, *U. urealyticum*, AE002101, *Bacillus subtilis*, D26185, *Bacillus halodurans*, AP001507, *Escherichia coli*, AB001340 and *Buchnera* sp. APS, AP001119.

**Southern blotting.** The *tmk-a* probe DNA was amplified by PCR using primers *tmka-N* (5'-TTG AAT TCC ATA TGA AAT TAA TCG TTT TTG AAG GAC T-3') and *tmka-C* (5'-TGA GCT CGA GTT AGT TAT GAT CGC CAT TTG ATA GTA CT-3'). The *tmk-b* probe DNA was amplified by PCR using primers *tmkb-N* (5'-TTG AAT TCC ATA TGT TTA TTT CTT TTG AAG GTT GTG A-3') and *tmkb-C* (5'-TGA GCT CGA GCT ATT TGA AAG ACT TCT TTG AGT TTT GT-3'). PCR amplification was performed in a thermal cycler (Perkin Elmer model 9700) using 30 cycles of denaturation for 15 s at 94 °C, annealing for 30 s at 55 °C, and extension for 1 min at 60 °C. DNA probes were labelled with glutaraldehyde, and hybridization and detection were carried out using the ECL Direct nucleic acid labelling and detection system (Amersham Biosciences). Purified OY phytoplasma genomic DNA and control DNA preparations from healthy plants were completely digested with *Hind*III, *EcoRV* or *Xba*I. Fragments were separated by agarose gel electrophoresis, and transferred to a positively charged nylon membrane (Boehringer Mannheim) (Sambrook *et al.*, 1989). The membrane was prehybridized in hybridization buffer [5 × SSC, 2% (w/v) blocking reagent, 0.1% (w/v) *N*-lauroylsarcosine, 7% (w/v) SDS, 50 mM sodium phosphate buffer (pH 7.0), 50% (v/v) formamide] at 42 °C for 1 h. Hybridization was carried out at 42 °C overnight in hybridization buffer containing the denatured probe. After hybridization, the membrane was washed twice in primary wash buffer (0.5 × SSC, 6 M urea and 0.4% SDS) at 42 °C for 20 min each, and then washed twice in secondary wash buffer (2 × SSC) at room temperature for 5 min each. Hybridized probes were detected following the manufacturer's guidelines.

**PCR amplification and cloning of *tmk-a* homologues.** In order to amplify the *tmk-a* homologues, we used primers *tmka-aF* (5'-ATG AAA TTA ATC GTT TTT GA-3') and *tmka-aR* (5'-TGA GCT CGA GGT TAT GAT CG-3') and performed PCR using OY-W total DNA as a template. PCR amplification was performed under the same conditions as described above. The amplified DNA fragment was cloned in the pGEM-T plasmid vector (Promega) and several cloned DNA fragments were sequenced.

**Expression of the phytoplasma *tmk* genes.** The pET system (Novagen) was used to generate a polyhistidine (polyHis)-tagged TMK fusion protein. We amplified *tmk-a* and *tmk-b* gene fragments by PCR using primers *tmka-N* and *tmka-C*, and *tmkb-N* and *tmkb-C*, respectively. The amplicon was then digested with *Nde*I and *Xho*I and inserted into pET30a. The polyHis-tagged TMKs were expressed in Epicurian Coli BL21-CodonPlus (DE3)-RIL (Stratagene). Cell extracts were applied to a nickel NTA-column (Novagen), washed with TBS buffer (20 mM Tris/HCl, pH 7.9 and 500 mM NaCl), and the fusion protein was eluted with TBS buffer containing 1 M imidazole. The purity of the protein was checked by SDS-PAGE and staining by Coomassie brilliant blue.

**Measuring TMK activity.** The kinase activities of the purified polyHis-tagged TMKs were measured by the method of Berghauer (1975), using an enzyme-coupled assay involving the following reactions:



**Fig. 1.** Schematic restriction endonuclease cleavage maps of the phytoplasma clones. The cloned regions are depicted by thick lines, and the restriction endonuclease cleavage sites are indicated. The open boxes show each ORF and the names of the predicted genes are indicated beside the boxes. Hatched boxes indicate the probes used for the Southern blot analysis shown in Fig. 3.

TMP + ATP → TDP + ADP (TMK)

Phosphoenol pyruvate (PEP) + ADP → pyruvate + ATP (pyruvate kinase)

Pyruvate + NADH → L-lactic acid + NAD<sup>+</sup> (lactate dehydrogenase)

A 10 µl sample of partially purified polyHis-tagged TMKs (TMK-a, 1.7 mg ml<sup>-1</sup>; TMK-b, 1.45 mg ml<sup>-1</sup>) was added to 90 µl reaction solution [100 µM TMP, 2 mM MgCl<sub>2</sub>, 100 mM Tris/HCl (pH 8.0), 200 µM PEP, 1.2 mM ATP, 80 µM KCl, 400 µM NADH, 1.5 U pyruvate kinase (Nacalai Tesque) and 1.5 U lactate dehydrogenase (TOYOBO)]. The reaction was allowed to proceed at room temperature for 60 min (one reading at zero time, and the second and later readings at 5 min intervals) and the consumption of NADH was determined by measuring the OD<sub>340</sub>. The production of 1 nmol ADP min<sup>-1</sup> was defined as 1 unit (U) of activity of TMK. The amount of purified protein was measured using a Bio-Rad protein assay kit. The lower limit of detectable activity was 0.1 U mg<sup>-1</sup>. For the inactivation of this fraction, partially purified TMK-a and TMK-b fractions were incubated at 98 °C for 5 min.

## RESULTS

### Cloning and DNA sequence of the *tmk* genes of OY-W phytoplasma

Total DNA was extracted from the OY-W phytoplasma-enriched fraction of infected garland chrysanthemum plants, digested, and cloned into pUC18 to construct a genomic library. The insert DNA from these clones was completely sequenced on both strands. We obtained two clones, WH18 and WH114, which shared similarities with *tmk* genes. Each contained an independent putative *tmk* gene, the *tmk-a* and *tmk-b* genes, which encoded proteins of 208 and 207 amino acids, respectively (Fig. 1). The GenBank

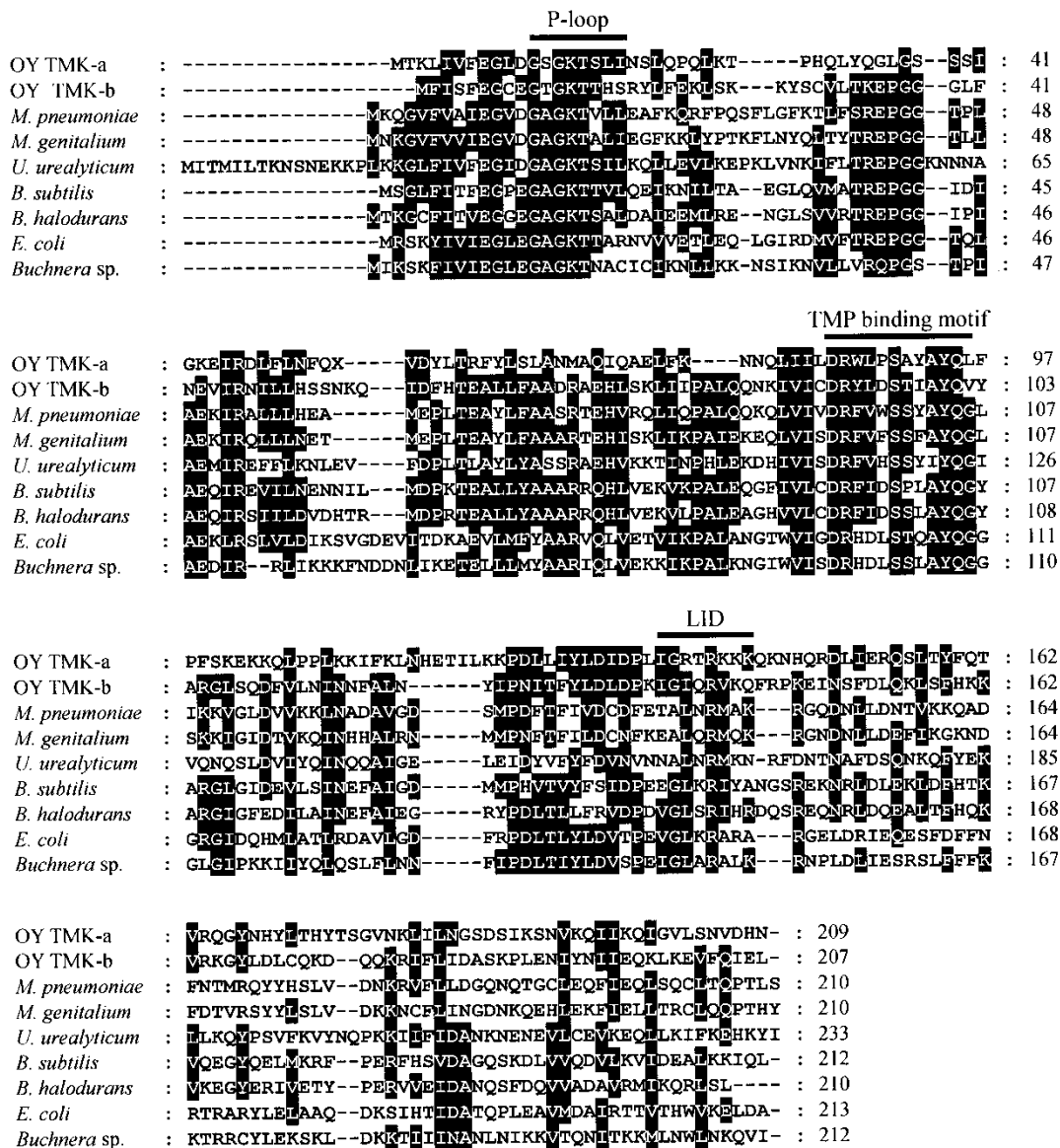
accession numbers of *tmk-a* and *tmk-b* are AB010446 and AB094668, respectively. Four ORFs with no similarity to any known functional genes were found upstream of *tmk-a* in WH18, while the *tmk-b* gene was encoded between the pseudouridylylase A gene (*truA*) and the DNA polymerase gamma subunit gene (*holB*) in WH114.

### *tmk* genes have conserved functional motifs

The deduced amino acid sequences of the proteins encoded by *tmk-a* (TMK-a) and *tmk-b* (TMK-b) had 43% identity. A comparison of TMK-a and TMK-b with predicted TMKs of other bacteria revealed that TMK-a shared 45% identity with *B. halodurans* TMK, and TMK-b shared 62% similarity with *B. subtilis* TMK (Table 1). The predicted amino acid sequences of seven bacterial TMKs were aligned with TMK-a and TMK-b (Fig. 2). Previous studies have reported that TMKs have specific motifs that are involved in NTP/NMP binding and that are conserved in other NMP kinases (Koonin & Senkevich, 1992; Traut, 1994; Lee & O'Brien, 1995; Reynes *et al.*, 1996; Lavie *et al.*, 1998). Fig. 2 shows that

**Table 1.** Amino acid sequence similarities (%) of TMKs

	TMK-a	TMK-b
<i>M. pneumoniae</i>	40	50
<i>M. genitalium</i>	40	50
<i>U. urealyticum</i>	39	46
<i>B. subtilis</i>	40	62
<i>B. halodurans</i>	45	59
<i>E. coli</i>	36	43
<i>Buchnera</i> sp.	40	43



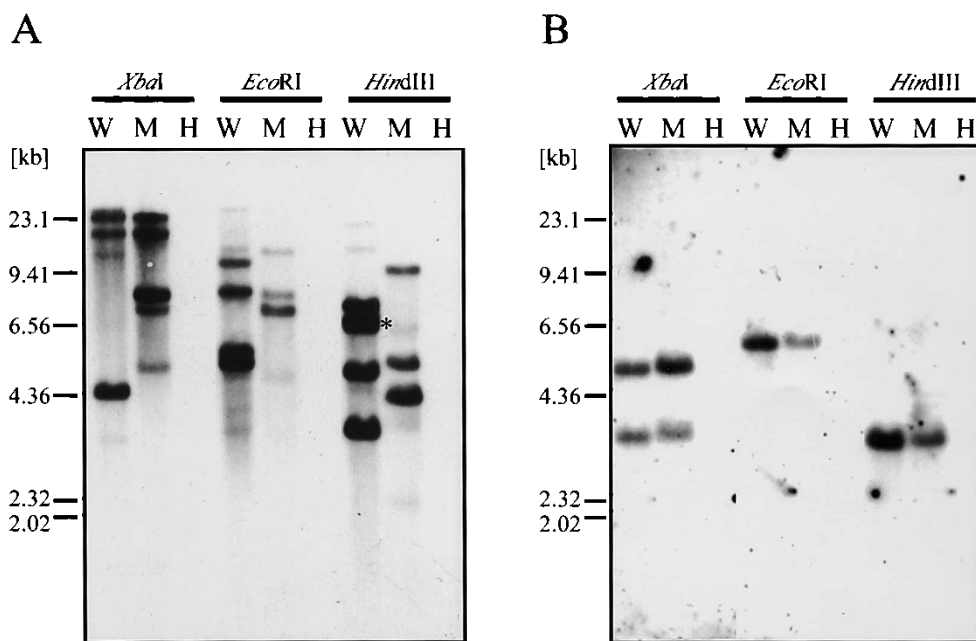
**Fig. 2.** Deduced amino acid sequence alignments. The deduced amino acid sequences of several bacterial TMKs and the OY-W TMKs were aligned using CLUSTAL W. Identical and similar residues are highlighted in black. The P-loop, TMP binding domains and LID are identified with thick lines above the sequence. OY TMK-a is the OY-W *tmk-a* product, and OY TMK-b is the OY-W *tmk-b* product.

the P-loop domain (N'-GXX GXX KT-C', where X indicates an arbitrary residue) is conserved in TMK-a (9-GLD GSG KT-16) and TMK-b (7-GCE GTG KT-14). The P-loop domain is involved in the binding of ATP and other phosphoryl donors. Furthermore, the TMP-binding motif N'-DRF XXS XXA YQ-C' was also conserved in both TMK-a (85-DRW LPS AYA YQ-95) and TMK-b (91-DRY LDS TIA YQ-101). Although the third motif residue in these sequences differed from that of the consensus motif, W (Trp), Y (Tyr) and F (Phe) are all aromatic amino acid residues. In addition, the LID region, which is another phosphoryl donor binding site, was also found. However, the amount of conservation was much lower, as has been reported

previously in humans, yeast and *E. coli* (Lavie *et al.*, 1998; Ostermann *et al.*, 2000). These results indicated that TMK-a and TMK-b have conserved functional motifs in their primary sequence, and suggest that they may function as TMK enzymes.

### Southern blot analysis of *tmk* genes

Total DNA extracted from OY-W- and OY-M-infected plants and from healthy plants was digested with *Hind*III, *Xba*I or *Eco*RI. The electrophoresed DNA fragments were blotted onto nylon membranes, and then hybridized with the *tmk-a* and *tmk-b* probes (Fig. 3). Signals were detected



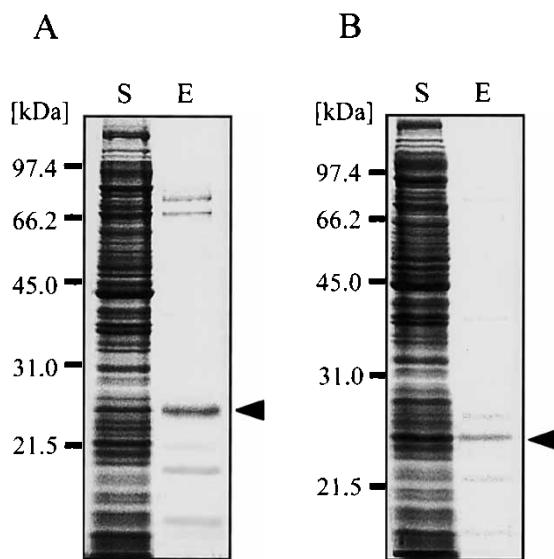
**Fig. 3.** Southern blot analysis of the *tmk-a* and *tmk-b* genes. Electrophoresed phytoplasma DNA isolated from OY-W- and OY-M-infected plants (lanes W and M), total DNA from a healthy garland chrysanthemum (lane H), and the restriction endonucleases used to digest the DNA, are indicated above each lane. Blotted DNA was hybridized with the *tmk-a* (A) or *tmk-b* (B) probes (shown in Fig. 1).  $\lambda$  DNA digested with *Hind*III was used as a molecular size marker. The asterisk in (A) indicates the signal corresponding to the WH18 clone.

in the lanes containing DNA from OY-W- and OY-M-infected plants, but not in the control lane containing healthy plant DNA. This strongly suggested that the signals originated from phytoplasma DNA. The *tmk-a* probe detected several fragments in each restriction endonuclease digest (Fig. 3A), while the *tmk-b* probe only detected one or two fragments (Fig. 3B). Although the coding region of *tmk-b* contained an *Xba*I site, none of the three endonucleases could digest the *tmk-a* gene in the WH18 clone. These results indicated that both *tmk-a* and *tmk-b* exist in the OY-W and OY-M genomes, that *tmk-b* is encoded as a single copy gene in both genomes, and that there may be several copies of *tmk-a*. The hybridization patterns in the OY-W lanes differed from those of OY-M (Fig. 3A).

In addition, we performed PCR using *tmk-a*F and *tmk-a*R primers in order to obtain *tmk-a* from the OY-W genome. Using OY-W total DNA as a template, a 0.6 kbp fragment was amplified. This amplified DNA was cloned and several clones sequenced. We obtained three *tmk-a* homologues, T01, T03 and T08, which shared 97, 93 and 95% identity, respectively, with *tmk-a*. However, because of point mutations, the ORF was not conserved in any of the clones (data not shown). Thus these homologues did not seem to be functional. The accession numbers of *tmk-a* homologues T01, T03 and T08 are AB100419, AB100420 and AB100421, respectively.

### ***In vitro* catalytic analysis of phytoplasma TMK proteins expressed in *E. coli***

In order to determine whether the *tmk-a* and *tmk-b* genes were functional, we expressed products of both genes using an *E. coli* expression system. Each *tmk* gene and a polyHis sequence were inserted 3' to the *E. coli* pET30 promoter. Each construct was used to transform *E. coli* cells, and the fusion proteins were overexpressed at 18 and 37 °C. The soluble fractions of the extracts of the *E. coli* cells were partially purified on Ni affinity columns. SDS-PAGE detected 25 kDa TMK-a-polyHis and TMK-b-polyHis fusion proteins (Fig. 4), which was consistent with their predicted molecular mass (TMK-a-polyHis, 25.0 kDa; TMK-b-polyHis, 24.9 kDa). To measure the TMK activity of these fractions, we used the methods described by Berghausen (1975). TMK activity was detected in the TMK-b-polyHis fraction ( $68.7 \pm 5.7$  U mg<sup>-1</sup>) and was destroyed ( $0.6 \pm 0.8$  U mg<sup>-1</sup>) by incubation of the fraction at 98 °C. No activity was detected in the TMK-a-polyHis fraction ( $0.2 \pm 0.6$  U mg<sup>-1</sup>). The negative control reactions without TMP (data not shown), or with elution buffer only ( $0.1 \pm 0.7$  U mg<sup>-1</sup>), had no activity. These results indicated that TMK-b has catalytic activity. We also tested a mixture of the TMK-a-polyHis and TMK-b-polyHis fractions and the catalytic activity of TMK-b was not affected (data not shown). This suggested that there was no inhibitory factor in the TMK-a-polyHis fraction.



**Fig. 4.** SDS-PAGE of fractions of *E. coli* cells overexpressing TMK-a-polyHis (A) or TMK-b-polyHis (B). Lanes S, sonicated soluble fraction; lanes E, partially purified eluates from the Ni column.

## DISCUSSION

We identified *tmk* gene homologues in the OY phytoplasma, which is an aster yellows (AY) 16S-group phytoplasma (Jung *et al.*, 2002). TMK is thought to be one of the most conserved enzymes and is found in viral, bacterial and eukaryotic genomes, but differences in activity might enable development of novel antimicrobial drugs to cure phytoplasma-infected plants.

The deduced amino acid sequences of OY phytoplasma TMK-a and TMK-b resembled those of *Bacillus* spp., *Mycoplasma* spp. and *U. urealyticum* more than those of Gram-negative bacteria (*E. coli* or *Buchnera* APS sp.) (Table 1, Fig. 2), which is consistent with the classification of the class *Mollicutes* in the *Bacillus/Clostridium* group. In addition, the *holB* gene, downstream of the *tmk* gene, which is conserved in several bacterial genomes including *M. genitalium*, *U. urealyticum*, *B. subtilis*, *Staphylococcus aureus* and *E. coli*, was also conserved in the *tmk-b* locus. Further investigation of the regions flanking *tmk-a* and its homologues provide further information about the origin of *tmk-a*.

The TMK-b-polyHis fusion protein had kinase activity, indicating that *tmk-b* encodes a functional TMK protein. However, the TMK-a-polyHis fusion protein had no catalytic activity. As all the functional motifs were conserved in TMK-a, it seems likely that TMK-a is functional, and experiments to detect its catalytic activity are being conducted. There may be several reasons why we could not detect catalytic activity for TMK-a. The optimal reaction conditions for TMK-a may differ from those of TMK-b. As

phytoplasmas inhabit both plant phloem sieve cells and insect cells, they may switch their metabolic pathways in response to change in habitat and TMK-a and TMK-b may function in independent pathways. As the same conserved motifs are also found in other NTP/NMP kinases, TMK-a might function as another type of kinase in phytoplasmas. For example, in *Mycoplasma* spp., deoxyadenosine kinase can phosphorylate not only deoxyadenosine but also deoxyguanosine and deoxycytidine (Wang *et al.*, 2001). Contaminating proteins that were not removed from the TMK-a-polyHis fraction may have inhibited the catalytic reactions in our experiments, or the fused polyHis-tag protein may have inhibited the reaction. Finally, the two highly conserved regions – TKEPGG, downstream of the P-loop motif, and PAL, upstream of the TMP binding motif – are absent in TMK-a (Fig. 2). There are no reports of the function of these motifs, but it is possible that the loss of the TKEPGG and PAL sequences impaired the TMK activity of TMK-a.

Southern blot hybridization analyses suggested that several homologous *tmk-a* sequences exist in both the OY-W and OY-M genomes, in addition to the one *tmk-b* gene. The *tmk* gene is largely encoded as a single copy gene in those organisms for which the genome has been completely determined (<http://www.ncbi.nlm.nih.gov/COG/index.html>). Two *tmk* gene homologues have been reported in the *B. subtilis* (BS-yorR), *Methanothermobacter thermoautotrophicus* (MTH1100) and *Archaeoglobus fulgidus* (AF1308) genomes. However, none of the additional homologues possess the TMP-binding motif, which is one of the three functional domains conserved in all known TMKs. This suggests that the OY phytoplasma is unique in encoding more than two copies of *tmk* genes containing catalytic domains.

Our findings suggest that gene duplication events have generated multiple copies of *tmk-a* in the phytoplasma genome. Ohno (1970) suggested that gene duplication was a major force in genome evolution and that large-scale gene duplication may enable the evolution of novel functions, an important step for the evolution of complex phenotypes. However, it is still unclear how often gene duplications arise and how frequently they evolve new functions. Indeed, there are only a few examples of the development of completely new functions (Cheng & Chen, 1999; Manzanares *et al.*, 2000). Thus it is generally thought that the fate of most duplicated genes is degeneration into pseudogenes. Lynch & Conery (2000) analysed the rate of gene duplication by examining several eukaryotic species and suggested that duplicate genes arise at a very high rate, on average 0.01 per gene per million years, but that most of the duplicates are silenced within a few million years.

Gene duplication has mainly been studied in eukaryotes. However, recently the complete genome sequences of several organisms, including prokaryotes, have been analysed to examine the relationships between gene function and the propensity of a gene to duplicate, and the number of genes in a gene family and the family's rate of sequence evolution (Conant & Wagner, 2002). For example, ribosomal genes

and transcriptional factors appear less likely to undergo gene duplication than other genes. In the case of the OY phytoplasma *tmk-a* homologues, some ORFs were destroyed by mutations. However, the *tmk-a* gene has a complete ORF and highly conserved functional domains, suggesting that TMK-a might have some other nucleotide kinase activity. Thus it is possible that some *tmk* genes have degenerated, and some have gained a new function. In mollicutes, there are some homologous genes in the COG database; there are two copies of the 50S ribosomal gene (MPN069 and 471) that encodes ribosomal protein L33 (RPL33) in *M. pneumoniae*. In the same organism MPN140 and 549 code for a 28 kDa protein that is part of the P1 operon. Large families of lipoprotein genes, such as the *vsp* genes of *Mycoplasma bovis* (Pfützner & Sachse, 1996), *rsa* genes of *M. pulmonis* (Simecka *et al.*, 1992) and *avg* genes of *M. agalactiae* (Flitman-Tene *et al.*, 2000), appear to play roles in adaptive systems for survival (Razin *et al.*, 1998). These surface lipoproteins are recognized by host immune systems, and variation in expression generates antigenic variation on the cell surface. However, TMK, RPL33 and other proteins encoded by duplicated genes may not play a role in host adaptation, and hence the reasons for maintenance of duplications of these genes may be different from those for lipoproteins. Thus the *tmk* genes of OY phytoplasma might be useful for further investigations of gene duplication in prokaryote genomes.

Southern blot analyses indicated that many *tmk-a* homologues are located in the OY phytoplasma genome and that the organization of these sequences may have been affected when OY-M was isolated from OY-W (Shiomi *et al.*, 1998). OY-M is a mildly pathogenic variant line in which about 130 kbp of the OY-W genome has been deleted (Oshima *et al.*, 2001). The regions flanking the *tmk-a* gene homologues may have been involved in the deletion.

Recently, several phytoplasma proteins were successfully expressed in *E. coli*, and used as antigens to obtain phytoplasma-specific antibodies. These proteins were unidentified membrane proteins (Yu *et al.*, 1998; Blomquist *et al.*, 2001), the SecA protein of the secretion system (Kakizawa *et al.*, 2001), and Rep proteins encoded in extrachromosomal DNA (Nishigawa *et al.*, 2001). These successes were based on the finding that phytoplasmas use the same universal codon system as *E. coli* and other bacteria for protein synthesis, unlike most other mycoplasmas (Lee *et al.*, 2000; Miyata *et al.*, 2002b). Similarly in this study we could express TMK-a-polyHis and TMK-b-polyHis fusion proteins in *E. coli*, and detected TMK-b catalytic activity. This is believed to be the first demonstration of catalytic activity of a phytoplasma protein using a heterologous expression system. Other phytoplasma proteins, such as many of enzymes involved in the metabolic pathways, may be obtained using the same approach. This information will be valuable for exploring new horizons in the study of the biological characteristics of phytoplasmas. Further analyses of the functions of phytoplasma-encoded proteins will answer many questions about the factors related to phytoplasma infection, plant symptoms,

host factors involved in phytoplasma infection, and ways to cure phytoplasma-infected plants by using novel antimicrobial therapeutics.

## ACKNOWLEDGEMENTS

This work was supported partly by Grants-in-Aid of Scientific Research from the Japan Society for the Promotion of Science (JSPS) (nos 09460155 and 13306004), a Grant-in-Aid of Scientific Research on Priority Areas (C) 'Genome Biology' from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Program for the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN) of the Bio-oriented Technology Research Advancement Institution.

## REFERENCES

- Altschul, S. F. & Lipman, D. J. (1990). Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* **87**, 5509–5513.
- Berghauer, J. (1975). A reactive arginine in adenylate kinase. *Biochim Biophys Acta* **397**, 370–376.
- Blomquist, C. L., Barbara, D. J., Davies, D. L., Clark, M. F. & Kirkpatrick, B. C. (2001). An immunodominant membrane protein gene from the Western X-disease phytoplasma is distinct from those of other phytoplasmas. *Microbiology* **147**, 571–580.
- Chambaud, I., Heilig, R., Ferris, S. & 9 other authors (2001). The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res* **29**, 2145–2153.
- Cheng, C.-H. & Chen, L. (1999). Evolution of an antifreeze glycoprotein. *Nature* **401**, 443–444.
- Conant, G. C. & Wagner, A. (2002). GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res* **30**, 3378–3386.
- Darby, G. K. (1995). In search of the perfect antiviral. *Antivir Chem Chemother* **6**, 54–63.
- Davies, D. L. & Clark, M. F. (1994). Maintenance of mycoplasma-like organisms occurring in *Pyrus* species by micropropagation and their elimination by tetracycline therapy. *Plant Pathol* **43**, 819–823.
- Doi, Y., Teranaka, M., Yora, K. & Asuyama, H. (1967). Mycoplasma or P.L.T. group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches'-broom, aster yellows or paulownia witches'-broom. *Ann Phytopathol Soc Jpn* **33**, 259–266.
- Flitman-Tene, R., Levisohn, S., Lysnyansky, I., Rapoport, E. & Yegorov, D. (2000). A chromosomal region of *Mycoplasma agalactiae* containing *vsp*-related genes undergoes *in vivo* rearrangement in naturally infected animals. *FEMS Microbiol Lett* **191**, 205–212.
- Fraser, C. M., Gocayne, J. D., White, O. & 26 other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
- Glass, J. I., Lefkowitz, E. J., Glass, J. S., Heiner, C. R., Chen, E. Y. & Cassell, G. H. (2000). The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* **407**, 757–762.
- Griffiths, P. D. (1995). Progress in the clinical management of herpesvirus infections. *Antivir Chem Chemother* **6**, 191–209.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C. & Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* **24**, 4420–4449.
- Jong, A. Y. & Campbell, J. L. (1984). Characterization of *Saccharomyces cerevisiae* thymidylate kinase, the CDC8 gene product.

- General properties, kinetic analysis, and subcellular localization. *J Biol Chem* **259**, 14394–14398.
- Jung, H.-Y., Sawayanagi, T., Kakizawa, T., Nishigawa, H., Miyata, S., Oshima, K., Ugaki, M., Joon-Tak, L. & Namba, S. (2002). 'Candidatus Phytoplasma castaneae', a novel phytoplasma taxon associated with chestnut witches' broom disease. *Int J Syst Evol Microbiol* **52**, 1543–1549.
- Kakizawa, S., Oshima, K., Kuboyama, T. & 7 other authors (2001). Cloning and expression analysis of phytoplasma protein translocation genes. *Mol Plant-Microbe Interact* **14**, 1043–1050.
- Kirkpatrick, B. C. (1992). Mycoplasma-like organisms: plant and invertebrate pathogens. In *The Prokaryotes*, 2nd edn. pp. 4050–4067. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K. H. Schleifer. New York: Springer.
- Koonin, E. V. & Senkevich, T. G. (1992). Evolution of thymidine and thymidylate kinases: the possibility of independent capture of TK genes by different groups of viruses. *Virus Genes* **6**, 187–196.
- Kuboyama, T., Huang, C. C., Lu, X., Sawayanagi, T., Kanazawa, T., Kagami, T., Matsuda, I., Tsuchizaki, T. & Namba, S. (1998). A plasmid isolated from phytopathogenic onion yellows phytoplasma and its heterogeneity in the pathogenic phytoplasma mutant. *Mol Plant-Microbe Interact* **11**, 1031–1037.
- Lavie, A., Ostermann, N., Brundiers, R., Goody, R. S., Reinstein, J., Konrad, M. & Schlichting, I. (1998). Structural basis for efficient phosphorylation of 3'-azidothymidine monophosphate by *Escherichia coli* thymidylate kinase. *Proc Natl Acad Sci U S A* **95**, 14045–14050.
- Lee, C. G. & O'Brien, W. E. (1995). A unique member of the thymidylate kinase family that is induced during macrophage activation. *J Immunol* **154**, 6094–6102.
- Lee, I.-M., Davis, R. E., Gundersen & Rindal, D. E. (2000). Phytoplasma: phytopathogenic mollicutes. *Annu Rev Microbiol* **54**, 221–255.
- Li de la Sierra, I., Munier-Lehmann, H., Gilles, A. M., Barzu, O. & Delarue, M. (2000). Crystallization and preliminary X-ray analysis of the thymidylate kinase from *Mycobacterium tuberculosis*. *Acta Crystallogr D Biol Crystallogr* **56**, 226–228.
- Lynch, M. & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.
- Manzanares, M., Wada, H., Itasaki, N., Trainor, P. A., Krumlauf, R. & Holland, P. W. (2000). Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* **408**, 854–857.
- McCoy, R. E., Caudwell, A., Chang, C. J. & 15 other authors (1989). Plant disease associated with mycoplasma-like organisms. In *The Mycoplasmas* vol. 5, pp. 546–640. Edited by R. F. Whitcomb & J. G. Tully. New York: Academic Press.
- Miyata, S., Furuiki, K., Sawayanagi, T., Oshima, K., Kuboyama, T., Tsuchizaki, T., Ugaki, M. & Namba, S. (2002a). The gene arrangement and sequence of *str* operon of phytoplasma resemble those of *Bacillus* more than those of *Mycoplasma*. *J Gen Plant Pathol* **68**, 62–67.
- Miyata, S., Furuiki, K., Oshima, K., Sawayanagi, T., Nishigawa, H., Kakizawa, S., Jung, H.-Y., Ugaki, M. & Namba, S. (2002b). Complete nucleotide sequence of the S10-*spc* operon of phytoplasma: gene organization and genetic code resemble those of *Bacillus subtilis*. *DNA. Cell Biol* **21**, 527–534.
- Nishigawa, H., Miyata, S., Oshima, K., Sawayanagi, T., Komoto, A., Kuboyama, T., Matsuda, I., Tsuchizaki, T. & Namba, S. (2001). *In planta* expression of a protein encoded by the extrachromosomal DNA of a phytoplasma and related to geminivirus replication proteins. *Microbiology* **147**, 507–513.
- Ohno, S. (1970). *Evolution by Gene Duplication*. New York: Springer-Verlag.
- Oshima, K., Shiomi, T., Kuboyama, T., Sawayanagi, T., Nishigawa, H., Kakizawa, S., Miyata, S., Ugaki, M. & Namba, S. (2001). Isolation and characterization of derivative lines of the onion yellows phytoplasma that do not cause stunting or phloem hyperplasia. *Phytopathology* **91**, 1024–1029.
- Ostermann, N., Schlichting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T., Goody, R. S. & Lavie, A. (2000). Insights into the phosphoryltransfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure* **8**, 629–642.
- Pfützner, H. & Sachse, K. (1996). *Mycoplasma bovis* as an agent of mastitis pneumoniae, arthritis and genital disorders in cattle. *Rev Sci Technol* **15**, 1477–1494.
- Pollack, J., Williams, M. & McElhaney, R. (1997). The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit Rev Microbiol* **23**, 269–354.
- Razin, S., Yogev, D. & Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**, 1094–1156.
- Reynes, J. P., Tiraby, M., Baron, M., Drocourt, D. & Tiraby, G. (1996). *Escherichia coli* thymidylate kinase: molecular cloning, nucleotide sequence, and genetic organization of the corresponding *tmk* locus. *J Bacteriol* **178**, 2804–2812.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Shiomi, T., Tanaka, M., Wakiya H. & Zenbayashi, R. (1996). Occurrence of Welsh onion yellows. *Ann Phytopathol Soc Jpn* **62**, 258–260.
- Shiomi, T., Tanaka, M., Sawayanagi, T., Yamamoto, S., Tsuchizaki, T. & Namba, S. (1998). A symptomatic mutant of onion yellows phytoplasma derived from a greenhouse-maintained isolate. *Ann Phytopathol Soc Jpn* **64**, 501–505.
- Simecka, J. W., Davis, J. K., Davidson, M. K., Ross, S. E., Stadtländer, C. T. K.-H. & Cassell, G. H. (1992). Mycoplasma diseases of animals. In *Mycoplasmas: Molecular Biology and Pathogenesis*, pp. 391–416. Edited by J. Maniloff, R. N. MacElhaney, L. R. Finch & J. B. Baseman. Washington, DC: American Society for Microbiology.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Traut, T. W. (1994). The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur J Biochem* **222**, 9–19.
- Wang, L., Westberg, J., Bolske, G. & Eriksson, S. (2001). Novel deoxynucleoside-phosphorylating enzymes in mycoplasmas: evidence for efficient utilization of deoxynucleosides. *Mol Microbiol* **42**, 1065–1073.
- Yu, Y. L., Yeh, K. W. & Lin, C. P. (1998). An antigenic protein gene of a phytoplasma associated with sweet potato witches' broom. *Microbiology* **144**, 1257–1262.