The Genome Reduction Excludes the Ribosomal Rescue System in Acholeplasmataceae

Christina Zübert\textsuperscript{a} Anna-Marie Ilic\textsuperscript{a} Bojan Duduk\textsuperscript{b} Michael Kube\textsuperscript{a}

\textsuperscript{a}Integrative Infection Biology Crops-Livestock, University of Hohenheim, Stuttgart, Germany; \textsuperscript{b}Institute of Pesticides and Environmental Protection, Belgrade, Serbia

Abstract

The \textit{trans}-translation process is a ribosomal rescue system for stalled ribosomes processing truncated mRNA. The genes \textit{ssrA} and \textit{smpB} fulfil the key functions in most bacteria, but some species have either lost these genes or the function of the ribosomal rescue system is taken over by other genes. To date, the ribosomal rescue system has not been analysed in detail for the \textit{Acholeplasmataceae}. This family, in the Mollicutes class, comprises the genus \textit{Acholeplasma} and the provisional taxon \textit{"Candidatus Phytoplasma"}. Despite their monophyletic origin, the two clades can be separated by traits such as not representing primary pathogens for acholeplasmas versus being phytopathogenic for the majority of phytoplasmas. Both taxa share reduced genomes, but only phytoplasma genomes are characterised by a remarkable level of instability and reduction. Despite the general relevance of the ribosomal rescue system, information is lacking on coding, the genomic context and pseudogenisation of \textit{smpB} and \textit{ssrA} and their possible application as a phylogenetic marker. Herein, we provide a comprehensive analysis of the ribosomal rescue system in members of \textit{Acholeplasmataceae}. The examined \textit{Acholeplasmataceae} genomes encode a ribosomal rescue system, which depends on tmRNA encoded by \textit{ssrA} acting in combination with its binding protein \textit{SmpB}. Conserved gene synteny is evident for \textit{smpB}, while \textit{ssrA} shows a less conserved genomic context. Analysis of the tmRNA sequences highlights the variability of proteolysis tag sequences and short conserved sites at the 5'- and 3'-ends. Analyses of \textit{smpB} provided no hints regarding the coding of pseudogenes, but they did suggest its application as a phylogenetic marker of \textit{Acholeplasmataceae} – in accordance with 16S rDNA topology. Sequence variability of \textit{smpB} provides sufficient information for species assignment and phylogenetic analysis.

Introduction

The ribosomal rescue system in bacteria enables the release of ribosomes that have stalled due to processed truncated mRNA. Such mRNAs are a result of the premature termination of gene transcription and/or endo- or...
exonucleolytic cleavage events. The rescue system has been reviewed in terms of its importance [Keiler, 2008; Janssen and Hayes, 2012; Himeno et al., 2014; Fritze et al., 2020]. It comprises a tmRNA and an SmpB protein. The small tmRNA is encoded by the ssrA gene (∼360 nt), which was formerly referred to as “105a-RNA” [Karzai et al., 1999], and has been described in a large number of bacteria after the first description in Escherichia coli [Ray and Apirion, 1979]. It is separated from structural RNAs by harbouring the tRNA and mRNA features elucidating the two encoded major functions – also reflected in the name “tmRNA” [Madigan et al., 2020]. An alanine-tRNA-like part, carrying an acceptor stem and a T-arm, enables the alanine load via the corresponding tRNA synthetase and by entering the A-site of the held ribosome, while the mRNA region of the tmRNA is located at the P-site [Komine et al., 1994]. As a result, the tmRNA-mediated trans-translation of the unblocked ribosome commences. The produced polypeptide is tagged by the tmRNA-encoded terminal peptide part AANDENYALAA (acting as a proteolysis tag peptide), recognised as the template for C-terminal-acting proteases in Es. coli [Tu et al., 1995; Withey and Friedman, 1999], which degrade the incomplete peptide chains [Keiler et al., 1996]. This process requires an essential accessory protein alongside the tmRNA, named “Smpb” for small protein B (e.g., ∼160 aa Es. coli K12) [Karzai et al., 1999; Shimizu and Ueda, 2002]. SmpB increases the alanylation activity of tmRNA, but it is also indispensable in the addition of tag-peptides by enabling the ribosomal A-site binding of tmRNA [Karzai et al., 1999; Barends et al., 2001; Shimizu and Ueda, 2002]. Furthermore, it protects tmRNA from degradation by RNase R [Hong et al., 2005]. In order to fulfil its role, SmpB has a C-terminal tail and a β-barrel domain at its core, with an oligonucleotide-binding (OB) fold [Dong et al., 2002; Someya et al., 2003]. SmpB and tmRNA in combination are essential for the trans-translation system and are encoded by most bacteria [Ge and Karzai, 2009]. However, an understanding emerges that some bacteria, either a functional ssrA cannot be found or smpB has an apparently inactivating mutation [Hudson et al., 2014]. These examples comprise Candidatus Carsonella ruddii strain PC, a highly obligate endosymbiont of the psyllid Ctenarytaina eucalypti. It encodes rudiments of a tmRNA, and a SmpB lacking the central loop and the C-terminal α-helix, thereby indicating pseudogenisation [Hudson et al., 2014]. This loss highlights the extreme genome reduction process of Candidatus Carsonella ruddii, resulting in a genome size of 160 kb for strain Pv [Nakabachi et al., 2006]. There are also bacteria with frameshifted smpB, such as Corynebacterium pseudotuberculosis 31, Mycobacterium intracellulare MOTT-2, Clostridium difficile str. CF5 and the str. M120, Buchnera aphidicola strains BCc and TLW03, Pectobacterium carotovorum PCC21, Aggregatibacter actinomycetemcomitans ANH9381, Pseudomonas putida DOT-T1E, Simiduia aurigavorans SAI, Mycoplasma pneumoniae FH, Thermotoga maritima MSB8, Petrotroba mobilis SJ95 and bacteria with truncated (the Tremblaya princeps strains PCIT and PCVAL) or pseudogenised smpB (Hodgkinia cicadica TETUND1). This process is seen as part of an evolutionary adaptation, and it has been observed in particular for bacteria enabled for intracellular colonisation, often characterised by genome reduction [Merhej et al., 2009]. Bacteria in the Mollicutes class are characterised by this type of degenerative evolution from gram-positive ancestors [Woese et al., 1980]. However, SmpB has been suggested as being part of the core gene set of minimal genomes [Mushegian and Koonin, 1996; Gil et al., 2004; Glass et al., 2006], and it has been shown to be preserved in species such as Mycoplasma genitalium [Fraser et al., 1995], which was used, due to its simplicity and small gene set, as a model for constructing the first synthetic cell. Despite the identification of the smpB gene in all Mollicutes [Grosjean et al., 2014], the highly adapted haemotrophic mycoplasmas lost the central loop region of SmpB [Hudson et al., 2014]. In addition, tmRNA seems to be absent from the Mycoplasma suis-subclade [Hudson et al., 2014]. The latter has been shown for Mycoplasma haemolamae Purdue, Mycoplasma suis Illinois, Mycoplasma wenyonii Massachusetts and Mycoplasma suis K13806. No alternative ribosome rescue system, such as ArfA/ArfB, seems to be encoded in the small genomes of these highly adapted Mollicutes [Grosjean et al., 2014].

In contrast to other major branches of Mollicutes, Acholeplasmataceae have not yet been examined in this respect in detail. This monophyletic branch consists of the eponymous genus Acholeplasma and the provisional taxon “Candidatus Phytoplasma” [IRPCM, 2004]. Phytoplasmas are known as insect-transmitted bacteria associated with hundreds of plant diseases, including those affecting many important crops [reviewed by Bertaccini and Duduk, 2009; Maejima et al., 2014; Kumari et al., 2019]. Acholeplasmas can be found as saprophytes in a variety of habitats or as commensals of vertebrates, insects, or plants. Differences between phytoplasmas and acholeplasmas are obvious on the genome level. Complete phytoplasma genomes range from 576 kb (“Candidatus Phytoplasma asteris” M3, acc. no. CP015149.1) to 960 kb (“Candidatus Phytoplasma australiense” NZSb11,
Ribosomal Rescue in Acholeplasmataceae

Complete genomes from the Acholeplasmataceae encode the key genes ssrA and smpB of the ribosomal rescue system. These single-copy genes show no frameshifts, truncations or other indications of function loss. The size of ssrA varies from 338 nt to 507 nt, whilst the encoded peptide tag ranges from 11 aa to 37 aa. The length of smpB ranges from 441 nt to 528 nt. No additional proteins forming an alternative ribosomal rescue system, such as ArfA/ArfB described for E. coli [Chadani et al., 2010; Chadani et al., 2011], were identified.

Results

Complete genomes from the Acholeplasmataceae encode the key genes ssrA and smpB of the ribosomal rescue system. These single-copy genes show no frameshifts, truncations or other indications of function loss. The size of ssrA varies from 338 nt to 507 nt, whilst the encoded peptide tag ranges from 11 aa to 37 aa. The length of smpB ranges from 441 nt to 528 nt. No additional proteins forming an alternative ribosomal rescue system, such as ArfA/ArfB described for E. coli [Chadani et al., 2010; Chadani et al., 2011], were identified.

The Genetic Context of Ribosomal Rescue System Genes

The genes smpB and ssrA are not located in close proximity on the chromosomes. A conserved genomic context is given for ssrA in “Candidatus Phytoplasma australiense” and “Candidatus Phytoplasma asteris” (online suppl. Fig. S1a, b; see www.karger.com/doi/10.1159/000520450 for all online suppl. material) but not in acholeplasmas (online suppl. Fig. S1c). In 4 of the 7 acholeplasmas, ssrA is flanked downstream by an IS3 family transposase, indicating the possible instability of this region. A less heterogenic situation is present for smpB genes flanked in phytoplasmas by an inorganic pyrophosphatase encoding gene (ppa) and in most acholeplasmas by ribonuclease R (RNase R) encoding gene as shown in Figure 1, thus highlighting a conserved genus-specific context on one border. The 3′-5′ exoribonuclease RNAse R is involved in degrading non-stop and defective mRNAs – and thereby providing a functional partner with the SmpB- tmRNA system in the trans-translation process [Richards et al., 2006]. The analyses highlight the increasing loss of the conserved gene order corresponding to the decreasing relatedness of taxa, albeit without an impact on the gene integrity of smpB and ssrA.

tmRNA and Encoded Peptide Proteolysis Tag

The deduced tmRNA sequences analysed herein have a conserved 5′- (GGGG) and 3′-end (CCACCA), in accordance with E. coli [Komine et al., 1994; Ushida et al., 1994; Karzai et al., 2000]. Based on amino acid sequence similarity, mainly for the conserved C-terminus, it was possible to identify the proteolytic tag peptide sequence of the tmRNA coding region (online suppl. Fig. S2). The tmRNA of Acholeplasmataceae members has a conserved region, starting with AUU as shown in Figure 2. In all analysed tmRNA sequences, this triplet is part of an open reading frame. In the genome of 3 strains, i.e., Achole-
**Fig. 1.** Gene order of the conserved anchoring of smpB. Regions encode inorganic pyrophosphatase and hypothetical proteins in the phytoplasma region (Pa, Pb), except for E. purpurea witches' broom phytoplasma (Pc), which is bordered by a gene encoding Hsp20/alpha crystallin family protein instead of a hypothetical protein, and by ribonuclease R (rnr) and patatin-like phospholipase family proteins in acholeplasmas (Aa), except for A. palmae (Ab), which is anchored by a phosphatase instead of patatin. Arrows symbolise forward or reverse strand coding, and information on the deduced gene products is provided.

*Non-homologous conserved hypothetical proteins in the "Ca. P. asteris" and australiense group.**

*Non-homologous conserved hypothetical proteins in the "Ca. P. asteris" and australiense group.* **Non-homologous hypothetical proteins.**

plasma brassicae, Italian clover phyllody phytoplasma and “Ca. P. mali”, a stop codon is located close upstream. The terminator TAA marks the 3'-end of the coding region in genomes of the Acholeplasmataceae.

Within phytoplasmas, all “Ca. P. asteris”-related strains share the same conserved N-terminal nucleotide sequence (5’-ATA ACC GGA AAA-3’) in the encoded peptide proteolysis tag, which translates into the amino acid sequence ITGN. “Ca. P. australiense” (rp-A) and “Ca. P. australiense” NZSb11 share 5’-ATA ACT GGA AAA-3’, while “Ca. P. solani” has an ACC triplet in the second position (5’-ATA ACC GGA AAA-3’), like “Ca. P. asteris”-related strains available, although both variants result in an amino acid motif ITGK. *Echinacea purpurea* witches’ broom phytoplasma NCHU2014, Italian clover phyllody (both 5’-ATA AAC GGC AAT-3’) and “Ca. P. ziziphi” (5’-ATA AAT GGC AAT-3’) share an amino acid sequence (5’-ATA AAC GGC AAT-3’) in the encoded peptide proteolysis tag, which translates into the amino acid sequence (5’-ATA ACC GGA AAT-3’) in the encoded peptide sequence (online suppl. Fig. S3).

Within Acholeplasma species, the start of the peptide coding sequence is highly conserved (5’-ATA ACC GGA AAC-3’). In Acholeplasma palmae (5’-ATA ACC GGA AAC-3’) and Acholeplasma equifetale (5’-ATA TCC GGA AAC-3’), there is one base deviation. All acholeplasmas share a conserved amino acid motif at the N-terminus (ITGN), except for A. equifetale (ISGN). A. palmae, A. brassicae, A. axanthum and A. modium exhibit a C-terminal sequence of ¹AA (¹ = F/L), whereas the other Acholeplasma species share the consensus ¹A¹A (¹ = L/Y/F) (online suppl. Fig. S4).

For all Acholeplasmataceae member sequences, except for “Ca. P. mali”, there is a conserved N-terminal amino acid sequence of ¹G¹ (¹ = T/N/S; G = N/K). All of the C-terminal residues are rather uncharged and hydrophobic (L, A, F, V, Y), and they are preceded by a cluster of polar and hydrophilic amino acids (T, Q, N, S), which are found in other eubacterial ssrA encoded peptide sequences [Karzai et al., 2000].

**Application of smpB as a Phylogenetic Marker**

The smpB nucleotide sequences exhibited pairwise distances up to 49.4% (A. equifetale and “Ca. P. pruni”)
Ribosomal Rescue in *Acholeplasmataceae*

It has been established herein that the ribosomal rescue system is a core genetic feature among members of the *Acholeplasmataceae* despite the fact that the genomic context of the key genes *ssrA* and *smpB* is not well conserved. Hudson et al. [2014] described *ssrA* as one of the most frequent neighbours of *smpB*, together with *ratA* (RatA toxin-inhibiting 70S ribosome association), *rnfH* (RnfH of ubiquitin superfamily) and RNase R. None of those genes identified is in direct proximity with the *smpB* of *Acholeplasmataceae* members analysed in this study, except for RNase R.

Despite the softened gene order within the genera, no hints for the pseudogenisation of the key genes have been obtained. Both tmRNA and SmpB retain general motifs beside group-specific features, and the 3′-terminal CAA trinucleotide is described as typical for mature tRNAs and tmRNAs [Komine et al., 1994; Ushida et al., 1994; Karzai et al., 2000]. In *E. coli*, pre-tmRNA must be processed into a mature and functional tmRNA. Therefore, cleavage at the 5′-end by RNase P [Komine et al., 1994], and 3′-end trimming by exoribonucleases, is necessary [Li et al., 1998] in a process which resembles that of canonical tRNA. These conserved endings are important for 5′- and 3′-pairing in secondary structure formation [Zwieb et al., 1999]. Alignments indicate (online suppl. Fig. S2) that the 3′- and 5′-endings of tmRNA sequences are conserved, in contrast to the middle parts. The conserved amino acid start motif for tmRNA-mediated peptide tagging and proteolysis in members of the *Acholeplasmataceae* does not match the one identified in *E. coli* [Keiler et al., 1996] and other Mollicutes, such as *M. pneumoniae* (DKNNDENVLPMLIANQASINYFA) [Zwieb et al., 1999] and *M. genitalium* (DKENNEV-LVDPNLINQASVNFA) [Karzai et al., 2000]. This may indicate that *Acholeplasmataceae* members exhibit a distinct amino acid start motif. The situation differs for mycoplasma *ssrA* tag endings, which are important for recognition by proteases, showing the consensus sequence N@A@A (⊕ = F/Y/L) [Gur and Sauer, 2008]. For *Acholeplasmataceae* members, this motif can be found in *A. hippoclon* (NYALA), but it is less conserved (⊕@A@A (⊕ = F/Y/L) in other *Acholeplasmataceae* species (*A. equifetale*, *A. oculi, A. granularum, A. laidlawii*) and in “Ca. P. aus-

(online suppl. Fig. S5). The distances within *Acholeplasma* species ranged from 21.1% (*Acholeplasma laidlawii* and *Acholeplasma granularum*) to 44.8% (*A. equifetale* and *A. brassicae*). Distances between “*Candidatus Phytoplasma*” clusters ranged from 19.2% (“Ca. P. australiense” NZSb11 and “Ca. P. asteris” OY-V, M3 and DY2014) to 34.7% (“Ca. P. pruni” CX and “Ca. P. australiense” rp-A), whereas the distances within clusters dropped, while some also exhibited identical sequences.

The high number of deviations in the tmRNA sequences in members of the *Acholeplasmataceae* disqualified them for application in phylogenetic analyses (data not shown), in contrast to *smpB*, which enables the evolutionary reconstruction of species and in some cases strain differentiation (shown in Fig. 3). Furthermore, the phylogenetic reconstruction also agreed with 16S rDNA analysis (online suppl. Fig. S6).

**Discussion**

It has been established herein that the ribosomal rescue system is a core genetic feature among members of the *Acholeplasmataceae* despite the fact that the genomic context of the key genes *ssrA* and *smpB* is not well conserved. Hudson et al. [2014] described *ssrA* as one of the most frequent neighbours of *smpB*, together with *ratA* (RatA toxin-inhibiting 70S ribosome association), *rnfH* (RnfH of ubiquitin superfamily) and RNase R. None of those genes identified is in direct proximity with the *smpB* of *Acholeplasmataceae* members analysed in this study, except for RNase R.

Despite the softened gene order within the genera, no hints for the pseudogenisation of the key genes have been obtained. Both tmRNA and SmpB retain general motifs beside group-specific features, and the 3′-terminal CAA trinucleotide is described as typical for mature tRNAs and tmRNAs [Komine et al., 1994; Ushida et al., 1994; Karzai et al., 2000]. In *E. coli*, pre-tmRNA must be processed into a mature and functional tmRNA. Therefore, cleavage at the 5′-end by RNase P [Komine et al., 1994], and 3′-end trimming by exoribonucleases, is necessary [Li et al., 1998] in a process which resembles that of canonical tRNA. These conserved endings are important for 5′- and 3′-pairing in secondary structure formation [Zwieb et al., 1999]. Alignments indicate (online suppl. Fig. S2) that the 3′- and 5′-endings of tmRNA sequences are conserved, in contrast to the middle parts. The conserved amino acid start motif for tmRNA-mediated peptide tagging and proteolysis in members of the *Acholeplasmataceae* does not match the one identified in *E. coli* [Keiler et al., 1996] and other Mollicutes, such as *M. pneumoniae* (DKNNDENVLPMLIANQASINYFA) [Zwieb et al., 1999] and *M. genitalium* (DKENNEV-LVDPNLINQASVNFA) [Karzai et al., 2000]. This may indicate that *Acholeplasmataceae* members exhibit a distinct amino acid start motif. The situation differs for mycoplasma *ssrA* tag endings, which are important for recognition by proteases, showing the consensus sequence N@A@A (⊕ = F/Y/L) [Gur and Sauer, 2008]. For *Acholeplasmataceae* members, this motif can be found in *A. hippoclon* (NYALA), but it is less conserved (⊕@A@A (⊕ = F/Y/L) in other *Acholeplasmataceae* species (*A. equifetale*, *A. oculi, A. granularum, A. laidlawii*) and in “Ca. P. aus-
Candidatus Phytoplasma cynodontis
Mycoplasma genitalium
Echinacea purpurea witches’ broom phytoplasm
'Candidatus Phytoplasma aurantifolia' WBDL
'Candidatus Phytoplasma solani' SA-1
'Candidatus Phytoplasma australiense' NZSb11
'Candidatus Phytoplasma australiense' (p-A)
'Candidatus Phytoplasma mail' AT
Peanut witches’ broom phytoplasma NTU2011
Echinacea purpurea witches’ broom phytoplasm NCHU2014
'Candidatus Phytoplasma pruni' CX
Italian clover phyllody phytoplasma MA1
'Candidatus Phytoplasma phoenicium' SA213
'Candidatus Phytoplasma pini' MDPP
'Candidatus Phytoplasma ziziphi' Jwb-nky
'Candidatus Phytoplasma cynodontis' LW01
'Candidatus Phytoplasma sacchari' SCGS 27
'Candidatus Phytoplasma oryzae' Mbila1
Acholeplasma palmae J233
Acholeplasma brassicae O502
Acholeplasma axanthum NCTC10138
Acholeplasma modicum PG 49
Acholeplasma hijikoroi NCTC10172
Acholeplasma equifetalae C112
Acholeplasma oculi 19L
Acholeplasma granularum BTS-39
Acholeplasma laidlawii PG-8A
Acholeplasma laidlawii NCTC10116
Mycoplasma genitalium G37
Ribosomal Rescue in Acholeplasmataceae

Fig. 3. Phylogenetic analysis of smpB in Acholeplasmataceae. The tree is inferred from smpB nucleotide sequences, using the Maximum Likelihood method and the Tamura-Nei model [Tamura and Nei, 1993]. In total, there are 574 positions in the final dataset. The bootstrapped confidence interval is based on 1,000 replications, and bootstrap values over 70% are shown on the branches. Evolutionary analyses were conducted in MEGA X [Kumar et al., 2018]. M. genitalium is used as an outgroup.

Ribosomal Rescue in Acholeplasmataceae
ITGN motive is also shared with the “Ca. P. asteris” strains but differs from other “Candidatus Phytoplasma” species. In contrast, smpB has a genus-specific genomic context and is applicable for the phylogenetic analysis of Acholeplasmataceae, thereby suggesting it as a marker gene in follow-up studies.

**Materials and Methods**

**Sequences**

SmpB protein sequences were retrieved from the Universal Protein Resource (uniprot.org) for Acholeplasmataceae with the taxonomy ID 2146 and a HAMAP family profile MF_00023. Additional sequences were identified by BLASTP, against NCBI’s non-redundant protein database (www.ncbi.nlm.nih.gov). Thirty-seven SmpB sequences were selected (online suppl. Table S1), belonging to the Interpro SsrA-binding protein family (IPR000037) and the CDD conserved protein domain family SmpB (cd09294), containing information about residues and positions at the SmpB-tmRNA interface. Based on protein IDs, nucleotide sequences were downloaded from the NCBI’s nucleotide database (Table 1).

**Table 1. Accession numbers of genome sequences and smpB base range**

| Taxon | Acc. No., base range | Length, nt |
|-------|----------------------|------------|
| Acholeplasma axanthum NCTC 10138 | LR215048.1, 1085411-1085857 | 447 |
| Acholeplasma brassicae OS02 | FO681348.1, c1130183-1129734 | 450 |
| Acholeplasma hippocion NCTC10172 | LR215050.1, c627356-626886 | 471 |
| Acholeplasma laidlawii PG-8A | CP000896.1, 427129-427578 | 450 |
| Acholeplasma laidlawii NCTC10116 | LS483439.1, 425709-426158 | 450 |
| Acholeplasma occil 19L | LK028559.1, c1121419-1120949 | 471 |
| Acholeplasma palmae J233 | FO681347.1, c1095332-1094889 | 444 |
| Acholeplasma modicum PG 49 (ATCC 29102) | NZ_JHYB010000002.1, 152088-152531 | 450 |
| Acholeplasma granularum BTS-39 (ATCC 19168) | NZ_JAFR010000006.1, 194990-195439 | 450 |
| Acholeplasma equifetale C1 12 (ATCC 29724) | NZ_JHXL010000009.1, 11661-12110 | 471 |
| “Candidatus Phytoplasma asteris” LD1 | MIEP010000004.1, 27059-27250 | 462 |
| “Candidatus Phytoplasma asteris” De Villa | CP035949.1, c501449-500988 | 462 |
| “Candidatus Phytoplasma asteris” OY-V | BBIY010000025.1, c10951-10225 | 459 |
| “Candidatus Phytoplasma asteris” DY2014 | SRMC010000000.1, c176675-176217 | 459 |
| “Candidatus Phytoplasma asteris” M3 | CP015149.1, c478812-478354 | 459 |
| “Candidatus Phytoplasma asteris” TW1 | QGKT010000002.1, 320594-321085 | 459 |
| “Candidatus Phytoplasma asteris” OY-M | AP006628.2, 172267-172725 | 459 |
| “Candidatus Phytoplasma asteris” RP166 | CP055264.1, 263909-264367 | 459 |
| “Candidatus Phytoplasma asteris” CYP | NZ_JSWN010000009.1, 3074-3532 | 459 |
| “Candidatus Phytoplasma tritici” | NZ_AVA001000001.1, 7784-78245 | 462 |
| “Candidatus Phytoplasma asteris” AYWB | CP0000001.1, c595870-595412 | 459 |
| “Candidatus Phytoplasma asteris” SA-1 | MPBG010000008.1, 27632-28084 | 459 |
| “Candidatus Phytoplasma asteris” NZSb11 | CP002548.1, 33958-340098 | 531 |
| “Candidatus Phytoplasma asteris” NZSb11 | AM422018.1, 596833-597363 | 531 |
| “Candidatus Phytoplasma asteris” AT | CZ469464.1, 596833-597363 | 531 |
| Peanut witches’ broom phytoplasma NTU2011 | AMWZ010000008.1, c21340-20891 | 456 |
| Echinacea purpurea witches’ broom phytoplasma NCHU2014 | CP040929.1, 528047-527598 | 450 |
| “Candidatus Phytoplasma aurantifolia” WBDL | MWKN010000047.1, 99727-99258 | 450 |
| “Candidatus Phytoplasma pruni” CX | LHC010000009.1, 3795-4247 | 453 |
| Italian clover phyllody phytoplasma MA1 | NZ_AKM010000028.1, c12328-11876 | 453 |
| “Candidatus Phytoplasma phoenicium” SA213 | JPSQ01000073.1, 2133-2528 | 450 |
| “Candidatus Phytoplasma pini” MDPP | VIE01000002.1, c63130-62684 | 447 |
| “Candidatus Phytoplasma ziziphi” Jwb-nky | CP025121.1, 431978-432424 | 447 |
| “Candidatus Phytoplasma cynodontis” LW01 | VW0H01000004.1, 2204-2674 | 471 |
| “Candidatus Phytoplasma sacchari” SSGS 27 | VWMK010000026.1, 2007-2465 | 459 |
| “Candidatus Phytoplasma oryzae” Mbita 1 | LTBM010000031.1, c9699-9238 | 459 |
| Mycoplasma genitalium G37 | L43967.2, c67643-67206 | 438 |
Ribosomal Rescue in Acholeplasmataceae

(CDK10644.1), “Ca. P. australiense” (rp-A) (CDK10643.1) and A. laidlawii PG-8A (CDK05533.1), using the Artemis genome browser [Carver et al., 2012]. The putative consensus start of a proteolysis tag peptide within the tmRNA sequence was defined by a conserved motif identified in nucleotide alignment, while the end was identified based on a conserved amino acid motif followed by a stop codon.

Encoded endopeptidase La (lon) was identified in genome entries (online suppl. Table S2). Protein families and domains were analysed with InterProScan [Jones et al., 2014].

Multiple sequence alignment of tmRNA and the proteolysis peptide tags was performed with CLUSTAL W (1.83), using T-COFFEE Version 11.00 [Notredame et al., 2000; Di Tommaso et al., 2011].

Phylogenetic Analyses

Nucleotide sequences of smpB gene were aligned by Clustal W, and a phylogenetic tree was reconstructed by using the Maximum Likelihood method and the Tamura-Nei model [Tamura and Nei, 1993] in MEGA X version 10.1.7 [Kumar et al., 2018]. A pairwise distance matrix was generated, using the p-distance method. All ambiguous positions were removed for each sequence pair (pairwise deletion option). In total, 573 positions were in the final dataset.

Table 2. Accession numbers of genome sequences and tmRNA locus tag or base range

| Taxon                        | Acc. No., locus tag or base range                  |
|------------------------------|---------------------------------------------------|
| Acholeplasma axanthum NCTC10138 | NZ_LR215048.1, EXC62_RS06055                      |
| Acholeplasma brassicaceae OS02 | NC_022549.1, BN853_RS08760                        |
| Acholeplasma hippikon NCTC10172 | NZ_LR215050.1, EXC59_RS02450                      |
| Acholeplasma laidlawii PG8     | NC_010163.1, ACL_RS07410                         |
| Acholeplasma laidlawii NCTC10116 | NL5483439.1, DN27_RS03880                       |
| Acholeplasma oculi 19L         | NL_LK028559.1, VV22_RS07795                       |
| Acholeplasma palmae J233      | NC_022538.1, BN854_RS07560                        |
| Acholeplasma modicum PG 49    | NC_JHYB01000002.1, T352_RS07010                   |
| Acholeplasma granularum BTS-39 | NC_JAFR01000006.1, G324_RS07825                   |
| Acholeplasma equifeutale C112 | NC_JHXL01000001.1, T434_RS07615                   |
| “Ca. P. australiense” (rp-A) | NC_007716.1, AYW_RB03535                         |
| “Ca. P. australiense” CYP      | NC_C035949.1, EXT02_RS05046                       |
| “Ca. P. australiense” De Villa | MIEPO1000001.1, BHEB2_00375                       |
| “Ca. P. australiense” M3       | NC_C015149.1, MBSPM3_RS00455                      |
| “Ca. P. australiense” NJAY     | MAFP0100001.1, BBA70_00330                       |
| “Ca. P. australiense” OY-M     | AP006628.2, 762465-762887                          |
| “Ca. P. australiense” OY-V     | QGTK01000002.1, DF280_02500                      |
| “Ca. P. australiense” RP166    | CP055264.1, 740426-740850                         |
| “Ca. P. australiense” TW1      | NC_010544.1, PAA_RS04365                         |
| “Ca. P. australiense” (rp-A)   | NC_022549.1, RL50_RS04365                         |
| “Ca. P. australiense” NZSb11   | NC_011047.1, c247166-247672                       |
| “Ca. P. australiense” mali AT  | NC_011047.1, c247166-247672                       |
| “Ca. P. australiense” solani SA-1 | NL_MPB01000007.1, SSSA1_RS02600                    |
| “Ca. P. australiense” triticci | NL_AVMO1000101.1, N747_RS03775                     |
| “Ca. P. australiense” Jwb-nky  | NL_CP025121.1, 577912-578413                      |

Phytoplasmas w/o detailed assignment in annotation

Echinacea purpurea witches’ broom NCUH2014 CP040925.1, 23588-23994
Italian clover phyllody phytoplasma str. MA1 NZ_AKIM0100013, RI01_RS03350

Phytoplasmas w/o detailed assignment in annotation

Echinacea purpurea witches’ broom NCUH2014 CP040925.1, 23588-23994
Italian clover phyllody phytoplasma str. MA1 NZ_AKIM0100013, RI01_RS03350

Table 3. Strains and GenBank entries used for examination of gene synteny (original submission and RefSeq annotation)

| Taxon                        | GenBank (acc. No./RefSeq entry) |
|------------------------------|---------------------------------|
| A. axanthum NCTC 10138       | LR215048.1/NZ_LR215048.1       |
| A. brassicaceae OS02          | FO681348.1/NC_022549.1        |
| A. hippikon NCTC10172        | LR215050.1/NZ_LR215050.1       |
| A. laidlawii PG-8A            | CP000896.1/NC_010163.1        |
| A. laidlawii NCTC10116        | NL5483439.1/NZ_L5483439.1     |
| A. oculi 19L                  | NL_LK028559.1/NZ_LK028559.1    |
| A. palmae J233               | FO681347.1/NC_022538.1        |
| “Ca. P. australiense” AYW     | CP000066.1/NC_007716.1        |
| “Ca. P. australiense” M3      | CP035949.1/NZ_CP035949.1      |
| “Ca. P. australiense” OY-M    | CP015149.1/NZ_CP015149.1      |
| “Ca. P. australiense” (rp-A)  | AP006628.2/–                   |
| “Ca. P. australiense” mali AT | AP006628.2/–                   |
| “Ca. P. australiense” Jwb-nky | AP006628.2/–                   |

Phytoplasmas w/o detailed assignment in annotation

Echinacea purpurea witches’ broom NCUH2014 CP040925.1, 23588-23994
Italian clover phyllody phytoplasma str. MA1 NZ_AKIM0100013, RI01_RS03350
The 16S rDNA sequences were retrieved manually from the GenBank genome entries (online suppl. Table S1). "Ca. P. asteris" OY-V (BB0Y00000000.1) lacks an annotated 16S rRNA gene entry. For "Ca. P. aurantifolia" and "Ca. P. oryzae", no smpB and 16S rDNA sequences were available from the same strain. M. genitalium G37 (L43967.2) was used as the outgroup, and smpB and 16S rDNA sequences were accessed from NCBI’s GenBank. Alignment was performed by Clustal W, and a phylogenetic tree was reconstructed by using the Maximum Likelihood method and the General Time Reversible model [Nei and Kumar, 2000] in MEGA X version 10.1.7 [Kumar et al., 2018]. The bootstrapped confidence interval was based on 1,000 replications.

All aligned matrices and trees were deposited in TreeBASE (www.treebase.org; accession number: 28778).

**Gene Synteny Analysis**

The gene context was analysed for ssrA and smpB from complete genomes (Table 3) in the Artemis genome browser [Carver et al., 2012]. Base ranges are shown in online supplementary Table S3. The genetic context of smpB was identified in INSDC entries, while ssrA was detected in RefSeq entries. For genomes in which ssrA was not annotated in the GenBank RefSeq genome entries, their position was inferred from their tmRNA nucleotide entries ("Ca. P. asteris" OY-M and "Ca. P. aurantifolia" RP166 and "Ca. P. oryzae", no smpB and 16S rRNA sequences were accessed from NCBI's GenBank. Alignment was performed by Clustal W, and a phylogenetic tree was reconstructed by using the Maximum Likelihood method and the General Time Reversible model [Nei and Kumar, 2000] in MEGA X version 10.1.7 [Kumar et al., 2018]. The bootstrapped confidence interval was based on 1,000 replications.

All aligned matrices and trees were deposited in TreeBASE (www.treebase.org; accession number: 28778).

**Statement of Ethics**

Ethical approval was not required as this research did not require any human and or animal involvement.

**Conflict of Interest Statement**

The authors have no conflicts of interest to declare.

**Funding Sources**

There were no funding sources.

**Author Contributions**

C.Z. and A.-M.I. carried out data collection, analysis and interpretation. C.Z., B.D. and M.K. performed the phylogenetic analysis and drafted the article.

All authors read and approved the final manuscript. All authors agreed to be both personally accountable for their own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.

**Data Availability Statement**

All data generated or analysed during this study are included in this article and its supplementary material files. The datasets for phylogenetic trees of Acholeplasmataceae (aligned matrices and trees) are deposited in TreeBASE (www.treebase.org/treebase, accession number: 28778).

**References**

Andersen MT, Liefting LW, Havukkala I, Beever RE. Comparison of the complete genome sequence of two closely related isolates of ‘Candidatus Phytoplasma australicen’ reveals genome plasticity. BMC Genomics. 2013;14: 529–15.

Bai X, Zhang J, Ewing A, Miller SA, Radek AJ, Shevchenko DV, et al. Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J Bacteriol. 2006;188:3682–96.

Barends S, Karzai AW, Sauer RT, Wower J, Krala B. Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl acceptor arm of tRNA. J Mol Biol. 2001;314:9–21.

Bertaccini A, Duduk B. Phytoplasma and phytosystema diseases: A review of recent research. Prototpathol Mediterr. 2009;48:355–78.

Bizarro CV, Schuck DC. Purine and pyrimidine nucleotide metabolism in Mollicutes. Genet Mol Biol. 2007;30(1 Suppl 1):190–201.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012;28:464–9.

Chadani Y, Ono K, Kutsukake K, Abo T. Escherichia coli YaeI protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. Mol Microbiol. 2011;80:772–85.

Chadani Y, Ono K, Ozawa SI, Takahashi Y, Takai K, Nanamiya H, et al. Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. Mol Microbiol. 2010;78:796–808.

de Novoa PG, Williams KP. The tmRNA website: Reductive evolution of tmRNA in plastids and other endosymbionts. Nucleic Acids Res. 2004;32:104–8.

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39:13–7.

Dong G, Nowakowski J, Hoffman DW. Structure of small protein B: The protein component of the tRNA-SmpB system for ribosome rescue. EMBO J. 2002;21:1845–54.

Duduk B, Bertaccini A. Phytoplasma classification: Taxonomy based on 16S ribosomal gene, is it enough? Phyt Moll. 2011;1(1):3–13.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, et al. The minimal gene complement of Mycoplasma genitalium. Science. 1995;270:397–403.

Fritze J, Zhang M, Luo Q, Lu X. An overview of the bacterial SsrA system modulating intracellular protein levels and activities. Appl Microbiol Biotechnol. 2020;104:5229–41.

Ge Z, Karzai AW. Co-evolution of multipartite interactions between an extended tmRNA tag and a robust Lon protease in Mycoplasma. Mol Microbiol. 2009;74:1083–99.

Gil R, Silva FJ, Peretó J, Moya A. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev. 2004;68:518–37.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, et al. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A. 2006;103:425–30.

Grojean H, Breton M, Sirand-Pugnet P, Tardy F, Thiacciourt F, Citti C, et al. Predicting the minimal translation apparatus: lessons from the reductive evolution of Mollicutes. PLoS Genet. 2014;10(5):e1004363.
Ribosomal Rescue in Acholeplasmataceae

Kumar S, Stecher G, Li M, Knysz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–57.

Kumari S, Nagendran K, Rai AB, Singh B, Govind Pratap Rao, Bertaccini A. Global status of phytoplasma diseases in vegetable crops. Front Microbiol. 2019;10:1–15.

Lazarov VN, Levitskii BA, Basovskii YI, Chukin MM, Akopian TA, Vereshchagin VV, et al. Complete genome and proteome of Acholeplasma laidlawii. J Bacteriol. 2011;193:4943–53.

Lee I-M, Gundersen SE, Davis RE, Bartoszky IM. Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. Int J Syst Bacteriol. 1998;48(4):1153–69.

Li Z, Pandit S, Deutscher MP. 3′ exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in Escherichia coli. Proc Natl Acad Sci U S A. 1998;95:2856–61.

Madigan MT, Benter KS, Buckley DH, Sattley MW, Stahl DA. Brock Mikrobiologie, ed 15. Pearson Deutschland GmbH; 2020.

Maejima K, Oshima K, Namba S. Exploring the phytoplasmas, plant pathogenic bacteria. J Gen Plant Pathol. 2014;80(3):210–21.

Manolakas JT, Barile MF, Chandler DK, Pollack JD. Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in Mollicutes. J Gen Microbiol. 1988;134:791–800.

Mao C, Driscoll T, Wower J, et al. Variations on the tmRNA gene. RNA Biol. 2009;6.

Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 2009;4:13–25.

Miles RJ. Catabolism in mollicutes. J Gen Microbiol. 1992;138:1773–83.

Mitrovic J, Kakizawa S, Duduk B, Oshima K, Namba S, Bertaccini A. The groEL gene as an additional marker for finer differentiation of the Acholeplasma laidlawii (subgroup A; rp-A) strains. Ann Appl Biol. 2011;159:41–8.

Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A. 1996;93:10268–73.

Nakabachi A, Yamashita A, Toh H, Ishihata K, Dunbar HE, Moran NA, et al. The 160-kilodalton genome of the plant-pathogenic phytoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.

Kameyama K, Miggold AM, et al. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics. 2008;9:1–14.
Ushida C, Himeno H, Watanabe T, Muto A. tRNA-like structures in 10Sa RNAs of Mycoplasma capricolum and Bacillus subtilis. Nucleic Acids Res. 1994;22:3392–6.

Wang J, Song L, Jiao Q, Yang S, Gao R, Lu X, et al. Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. BMC Genomics. 2018;19:1–12.

Wei W, Davis RE, Jomantiene R, Zhao Y. Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. Proc Natl Acad Sci U S A. 2008;105:11827–32.

Withey J, Friedman D. Analysis of the role of trans-translation in the requirement of tmRNA for λimm(P22) growth in Escherichia coli. J Bacteriol. 1999;181:2148–57.

Woese CR, Maniloff J, Zablen LB. Phylogenetic analysis of the mycoplasmas. Proc Natl Acad Sci U S A. 1980;77:494–8.

Zwieb C, Wower I, Wower J. Comparative sequence analysis of tmRNA. Nucleic Acids Res. 1999;27:2063–71.