

## Review Article

# Current View on Phytoplasma Genomes and Encoded Metabolism

Michael Kube,<sup>1,2</sup> Jelena Mitrovic,<sup>3</sup> Bojan Duduk,<sup>3</sup> Ralf Rabus,<sup>4,5</sup> and Erich Seemüller<sup>6</sup>

<sup>1</sup> Department of Crop and Animal Sciences, Humboldt-University of Berlin, Lentzeallee 55/57, 14195 Berlin, Germany

<sup>2</sup> Max Planck Institute for Molecular Genetics, Ihnestr. 63, 14195 Berlin, Germany

<sup>3</sup> Department of Plant Pathology, Institute of Pesticides and Environmental Protection, Banatska 31b, P.O. Box 163, 11080 Belgrade, Serbia

<sup>4</sup> Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University of Oldenburg, Carl-von-Ossietzky Straße 9-11, 26111 Oldenburg, Germany

<sup>5</sup> Department for Microbiology, MaxPlanck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

<sup>6</sup> Institute for Plant Protection in Fruit Crops and Viticulture, Federal Research Centre for Cultivated Plants, Schwabenheimer Straße 101, 69221 Dossenheim, Germany

Correspondence should be addressed to Michael Kube, michael.kube@agrar.hu-berlin.de

Received 26 October 2011; Accepted 20 November 2011

Academic Editors: M. J. Paul and T. Tanisaka

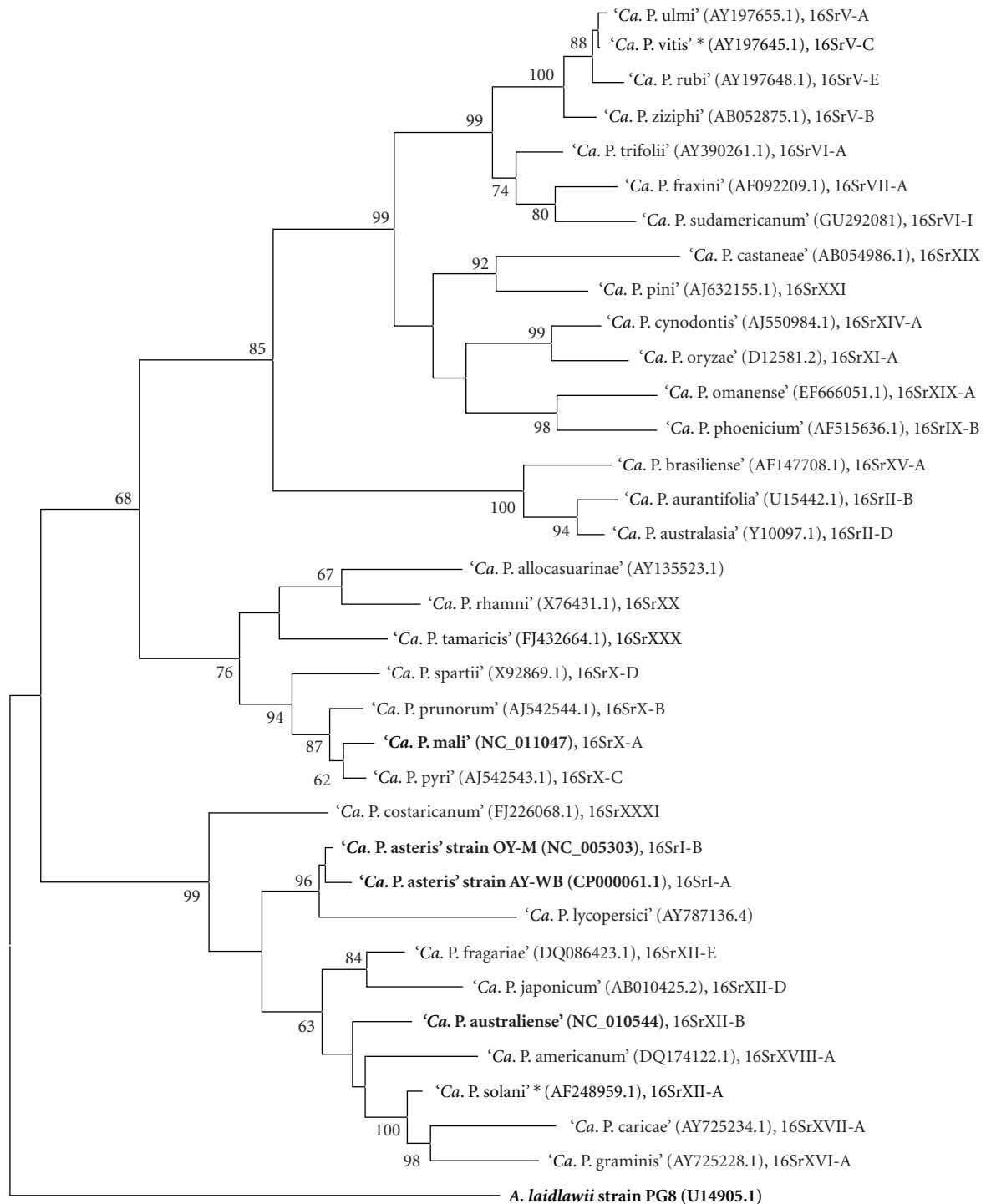
Copyright © 2012 Michael Kube et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Phytoplasmas are specialised bacteria that are obligate parasites of plant phloem tissue and insects. These bacteria have resisted all attempts of cell-free cultivation. Genome research is of particular importance to analyse the genetic endowment of such bacteria. Here we review the gene content of the four completely sequenced ‘*Candidatus* Phytoplasma’ genomes that include those of ‘*Ca. P. asteris*’ strains OY-M and AY-WB, ‘*Ca. P. australiense*’, and ‘*Ca. P. mali*’. These genomes are characterized by chromosome condensation resulting in sizes below 900 kb and a G + C content of less than 28%. Evolutionary adaption of the phytoplasmas to nutrient-rich environments resulted in losses of genetic modules and increased host dependency highlighted by the transport systems and limited metabolic repertoire. On the other hand, duplication and integration events enlarged the chromosomes and contribute to genome instability. Present differences in the content of membrane and secreted proteins reflect the host adaptation in the phytoplasma strains. General differences are obvious between different phylogenetic subgroups. ‘*Ca. P. mali*’ is separated from the other strains by its deviating chromosome organization, the genetic repertoire for recombination and excision repair of nucleotides or the loss of the complete energy-yielding part of the glycolysis. Apart from these differences, comparative analysis exemplified that all four phytoplasmas are likely to encode an alternative pathway to generate pyruvate and ATP.

## 1. Introduction

Phytoplasmas are bacteria which colonise plants and some groups of insects. They are associated with diseases of several hundred plant species including important crops [1]. Just for apple, economical losses caused by ‘*Candidatus* Phytoplasma mali’ and related phytoplasmas of 100 million €/year in Italy and Germany were estimated [2]. Phytoplasmas are pleomorphic but nonhelical, 0.2–0.8 μm in diameter, and lack, as members of the class *Mollicutes*, a firm cell wall. Phytoplasmas were first called mycoplasma-like organisms on discovery in 1967 [3]. Today, they are grouped in the provisional genus ‘*Candidatus* Phytoplasma’ that is assigned to a family of *incertae sedis* [4]. Thirty-two ‘*Candidatus* Phyto-

plasma’ species have been described [5–7]. The phytoplasma clade is divided into two major subclades and numerous subgroups based on the 16S rRNA marker gene (Figure 1). Based on their 16S rDNA sequence they are phylogenetically assigned to a distinct monophyletic taxon in the *Mollicutes* as member of the order *Acholeplasmatales* [8]. *Mollicutes* are deeply branching from the clostridial lineage leading to the genera *Bacillus* and *Lactobacillus* [9–11]. They are assigned to the phylum *Tenericutes*. Within the mollicutes, phytoplasmas are most closely related to the genus *Acholeplasma* [12, 13] with which they share, in contrast to other mollicutes, the usage of UGA as a stop codon [14]. Based on 16S rDNA analysis, Weisburg et al. [15] proposed that acholeplasmas were formed early at the initial divergence of the *Mollicutes*



20

FIGURE 1: Phylogenetic tree constructed by parsimony analyses of 16S rDNA sequences of 34 phytoplasma strains, belonging to all '*Candidatus Phytoplasma*' species described employing *A. laidlawii* as outgroup. The completely sequenced strains are in bold. CLUSTAL W from the Molecular Evolutionary Genetics Analysis program-MEGA5 [26] was used for multiple alignment. The maximum parsimony tree was obtained using the Close-Neighbour-Interchange algorithm, implemented in the MEGA5, with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). One of nine equally parsimonious trees is shown in Figure 1. Numbers on the branches are bootstrap values obtained for 2000 replicates (only values above 60% are shown). GenBank accession numbers are given in parentheses, and 16Sr group classification (when available) is shown on the right side of the tree. Asterisk indicates '*Candidatus Phytoplasma*' proposed by the IRPCM Phytoplasma/Spiroplasma Working Team—phytoplasma taxonomy group [5], not yet formally described. The tree is drawn to scale with branch lengths calculated using the average pathway method and represents the number of changes over the whole sequence. The scale bar represents 20 nucleotide substitutions.

from clostridial ancestors, followed by genome condensation and loss of the cell wall. Another divergence may have occurred that led to the separation of the phytoplasmas from the achleoplasmas. Genome condensation took place in both groups resulting in genomes sizes of about 1.215–2.095 kb and 27–38 mol% GC for the achleoplasmas [16, 17] and 530–1.350 kb and 21–33 mol% GC for the phytoplasmas [4, 17–20].

Information on the phytoplasma host environment is essential for understanding their adaptation and genome condensation. In plants, phytoplasmas mainly colonise the phloem sieve elements, but have occasionally also been reported to occur in companion and parenchyma cells as well [21–23]. The sieve tube environment is highly appropriate for the pathogens because it offers excellent conditions for the spread in the plant host. The phytoplasmas can readily pass the sieve pores of the sieve tubes and may be passively translocated according to the source/sink principle of the phloem sap flow in the various growing seasons [1, 24, 25].

The phloem fulfills functions in long-distance transport and allocation of nutrients and emits signals related to nutrition, development, and stress response. Probably the phloem transport functions by mass flow through the sieve elements and by cell-to-cell transport by surrounding companion cells and phloem parenchyma cells that provide metabolites to the sap [27].

Phloem sap is unique for several reasons. It is under high hydrostatic pressure and rich in nutrients. The content varies depending on the plant species but always contains large amounts of carbohydrates. The most abundant one is often sucrose (12–30%) beside other sugars such as glucose and fructose, minerals, proteins, amino acids, and ATP [28, 29]. Other sugars such as polyols and oligosaccharides of the raffinose family may be also abundant, depending on the species [30]. Sugar alcohols such as mannitol and sorbitol may also be present in considerable amount [31, 32].

The osmotic pressure of the sieve tube sap ranges from 0.8–2.5 MP and the pH is usually between 7.3 and 8.5 [30]. Bicarbonate and malate may play a role in the control of phloem sap pH [33, 34]. It is notable that malate and citrate represent the predominant organic acids in phloem and xylem sap [35], and one may speculate whether the content of these organic acids is regulated as a consequence of phytoplasma infection. It is known that the pH value of media used for cultivation of plant-derived isolates of *Achleoplasma brassicae* and *A. palmae* decreases significantly during growth.

Other nutrients such as nitrate, inorganic phosphate, glutathione, glutamate, and amino acids were also detected in phloem sap [36, 37]. For lettuce (*Lactuca sativa* L.) it was shown [28] that fourteen amino acids are present (Asp, Thr, Ser, Asn, Glu, Gln, Gly, Ala, Val, Ile, Leu, Tyr, Phe, and gamma-isobutyric acid/GABA). Glutamine, glutamic acid, serine, and GABA were often predominant. The concentration in lettuce was 125 mM for all amino acids. The presence of high ammonia concentrations (6 mM) is also remarkable. However, depending on the method also arginine and trace quantities of alpha-aminobutyric-acids, methionine, tryptophan, and ornithine were detected. Secondary sub-

stances such as phenolics, flavonoids, sesquiterpene lactones, alkaloids, and sterols have also been reported [28].

Some of the listed compounds and others such as cytokinins, auxins, abscisic acid, gibberellins, jasmonates, and methyl salicylates [27, 36, 37] and small RNAs [38, 39] highlight the importance of the phloem for signalling. It is still under discussion if the phloem is also a location for protein synthesis [27]. The determination of pumpkin sap proteome supports this idea [40].

Phytoplasmas are mainly transmitted by phloem-feeding insects such as leafhoppers, planthoppers, and psyllids [41, 42]. Under certain conditions, vectors may be attracted by plants releasing plant allomones manipulated by the phytoplasma infection [42]. They have to acquire the pathogen by phloem feeding [43, 44]. Subsequently, the phytoplasmas multiply in the hemolymph in which the concentration of organic compounds is similarly high than in phloem sap [45]. It is remarkable that trehalose is the predominant sugar in the hemolymph of aphids while glucose appeared as an artefact, and sucrose was not detected [45]. The trehalose content of the hemolymph of the phloem-sucking insects is correlated to the sucrose content of the aphid's host plants [45]. Interaction of phytoplasmas with their insect vectors has been reported for '*Ca. P. asteris*' strains. Dominant membrane proteins such as the major antigenic protein Amp interact with the vector proteins such as actin and the alpha and beta subunits of the ATP synthase [46]. The insect environment in combination with the ability to migrate from the insect gut to the salivary glands allows the spread of the pathogen [47].

Four complete phytoplasma genomes have been published. However, there is no recent publication summarizing the encoded repertoire of these genomes (see also ST, supplementary table in Supplementary Material available online at doi: 10.1100/2012/185942) and discussing the deduced metabolism. For this reason this paper attempts to fill this gap by reviewing published results and re-evaluating the genomic data.

## 2. Genomic Benchmarks

The first fully sequenced genome was that of the onion yellows phytoplasma strain OY-M of subgroup 16SrI-B [48]. It was followed by the closely related aster yellows witches'-broom phytoplasma strain AY-WB of subgroup 16SrI-A that showed large regions of conserved synteny [49]. In addition to the information on this two '*Ca. P. asteris*' genomes the sequences of two members from other phylogenetic groups were published in 2008, namely, that of '*Ca. P. australiense*' [50] and that of '*Ca. P. mali*' [20] (Table 1). With a size of 880 kb '*Ca. P. australiense*' shows the largest phytoplasma chromosome deciphered so far while '*Ca. P. mali*' with a size of 602 kb stands at the opposite end belonging to the second major subclade. '*Ca. P. mali*' is characterized by the extremely low G + C content of less than 22%, while the other three genomes show a similar G + C content of 27–28%. The number of protein-encoding genes ranges corresponding to their genome sizes from 481 to 776 but may be also influenced by different annotation styles. The

TABLE 1: Overview of the four complete phytoplasma genomes and *A. laidlawii*.

Strain	' <i>Candidatus</i> Phytoplasma' species					<i>Acholeplasma</i>
	asteris OY-M	asteris AY-WB	australiense Rp-A	mali AT	<i>laidlawii</i> PG-8A	
Chromosome organisation	Circular	Circular	Circular	Linear	Circular	
Chromosome size	853,092	706,569	879,959	601,943	1,496,992	
G + C content (%)	27.76	26.89	27.42	21.39	31.93	
G + C % of protein-coding genes <sup>1</sup>	29.09	28.54	28.72	22.58	32.23	
Protein-coding genes <sup>1,2</sup>	752	776	684 (155)	481 (16)	1,380 (11)	
Protein coding (%) <sup>1</sup>	73.1	73.7	64.1	76.3	90.7	
Average ORF size <sup>1</sup>	829	776	825	955	984	
Protein-coding genes/kb <sup>1</sup>	0.881	0.949	0.777	0.799	0.921	
rRNA operons	2	2	2	2	2	
tRNAs	32	31	35	32	34	
Accession no.	AP006628.2	CP000061.1	AM422018.1	CU469464.1	CP000896.1	
Plasmid-like elements	2	4	1	0	—	

<sup>1</sup>Without pseudo genes; <sup>2</sup>number of annotated pseudo genes is given in brackets.

highest number of strain-specific genes is present in '*Ca. P. australiense*' [50]. The only available genome sequence from the related genus *Acholeplasma* is that of *A. laidlawii* strain PG-8A (Acc. no. CP000896), which differs from phytoplasma genomes by a larger size (1.5 Mb), higher number of protein-encoding genes (1,380), and a higher G + C content (32%). However, all sequenced phytoplasmas belong to three subgroups leaving most subgroups out of this analyses (Figure 1).

The reason for the low number of completely sequenced phytoplasma genomes is not surprising if one considers the difficulties encountered. Mainly due to the problem in obtaining suitable DNA for sequencing, it took years to complete '*Ca. P. asteris*' strain OY-M and '*Ca. P. mali*' strain AT sequencing. Because phytoplasma DNA usually has to be extracted from infected plants in which it occurs at very low concentration, elaborate enrichment and purification procedures were required to obtain it in amounts of DNA necessary for a high coverage of each position of the genome in shotgun sequencing. Flowers of *Catharanthus roseus* (periwinkle) and phloem of greenhouse-grown *Nicotiana tabacum* (tobacco) plants were preferentially used to obtain DNA templates for '*Ca. P. australiense*' and '*Ca. P. mali*' sequencing, respectively. The amount of phytoplasma DNA obtained depends on the enrichment procedure used. For example, when '*Ca. P. mali*' DNA was, after extraction from *N. occidentalis*, treated by repeated bisbenzimidazole-CsCl buoyant density gradient centrifugation, a 30% enrichment of phytoplasma DNA was obtained. In contrast, purifying of phloem extract from *N. tabacum* using pulsed-field gel electrophoresis resulted in 80% enrichment [20]. Starting from an enriched phytoplasma DNA template, the genome sequences were determined by a shotgun approach [51]. In theory, shotgun fragments and the derived sequences represent the whole genome equally. However, AT-rich regions show a decreased physical coverage by recombinant clones/shotgun sequences and a lower sequence quality. Low coverage regions in the genomes are the result. This problem

also arises if new strategies from next generation technologies such as pyrosequencing [52] are used. The effect is reduced by these clone-independent approaches and high sequence coverage.

The assembly and finishing of the repeat-rich genomes represents an additional problem. It remains time consuming because clone-based large-insert libraries of phytoplasmas (e.g. fosmid libraries) do show a high cloning bias and are thus limited in their suitability for finishing experiments. The generation of additional complete genome sequence will remain a challenge in near future, but, hopefully, decreasing sequencing costs, high sequencing coverage, mate-pair libraries, and enlarged read lengths averaging 800 bases derived from pyrosequencing will help to overcome these problems.

The four phytoplasma genomes are organized in chromosomes and often contain extrachromosomal elements, which are reported for several phytoplasma strains [53]. The extrachromosomal elements, often called plasmids, allow the integration of their genetic material into the chromosome [49] and remarkably influence the vector transmissibility as shown for some strains [54]. Transmission experiments also indicate that plasmid-like extrachromosomal elements may not belong to the stable content of the genome.

The chromosomes of nearly all *Mollicutes* including those of the two '*Ca. P. asteris*' strains [48, 49], '*Ca. P. australiense*' [50], and *A. laidlawii* are circular. In contrast, '*Ca. P. mali*' and the closely related species '*Ca. P. pyri*' (pear decline phytoplasma) and '*Ca. P. prunorum*' (European stone fruit yellows phytoplasmas) have a linear chromosome [20]. It may indicate a characteristic of this phylogenetic cluster. The chromosome of '*Ca. P. mali*' has terminal inverted repeats (TIRs) with a size of 43 kb that have covalently closed hairpin ends protecting the chromosome [20]. TIRs are also known from *Streptomyces* genomes and covalently closed hairpin ends occur in genera *Borellia* and *Coxiella*, and several linear plasmids and phages [55]. A candidate gene encoding a required telomere resolvase (ResT) has been

suggested to occur in ‘*Ca. P. mali*’ chromosome (ATP\_00103) [20]. However, further experiments are needed to clarify the proposed function. Within the TIR region neither tRNAs nor rRNAs genes are encoded. The origin of this unusual chromosome organisation within the *Mollicutes* and its influence on the replication process in phytoplasmas remains unclear. One may speculate whether these structures have a virus-related origin.

Viruses influence the phytoplasma genomes. Wei and colleagues proposed that phages of the order *Caudovirales* have a major impact on the two fully sequenced ‘*Ca. P. asteris*’ genomes [56]. They analysed the deviations in the conserved synteny of these two closely related genomes and the encoded cryptic phage-derived genes and calculated that 264.2 kb (~31% of the genome) of strain OY-M and 160.2 kb (~22.7% of the genome) of strain AY-WB are putatively of viral origin. The Prophage Finder was one of the central bioinformatical tools used [56, 57]. Employing the Prophage Finder approach to examine the four fully sequenced phytoplasma genomes results in similar percentages of phage-derived sequences and shows 32% for strain OY-M, 20% for strain AY-WB, 26% for ‘*Ca. P. australiense*’ and 26% for ‘*Ca. P. mali*’. Even if such an approach can just give a raw estimation, it is obvious that these genomes are heavily influenced by integration events. Genome diversity of phytoplasmas was also shown in other studies, for example, by PFGE analysis and mapping of marker genes of the sweet potato little leaf (SPLL) disease strain V4 (SPLL-V4) and of the closely related tomato big bud (TBB) phytoplasma [58].

Genome instability of phytoplasmas is not restricted to phage integration and recombination events or exchangeable plasmid material. Complex transposons, called potential mobile units (PMUs) [49], have significant impact in duplication of parts of phytoplasma genomes. PMUs are suggested to act as a tool to generate genetic variability [49]. Recently, it was shown that they might form extrachromosomal elements that may replicate and integrate again into phytoplasma genomes [59]. A high portion of the encoded proteins in the four genomes was assigned to these elements. Hogenhout and Musić [60] assigned 486 (OY-M) to 408 (‘*Ca. P. mali*’) proteins of the four phytoplasma genomes as single copy genes. This means that 35% of all protein-encoding genes in OY-M are multicopy genes. The corresponding proteins encoded by multicopy genes were assigned to PMUs. This indicates that 65–70% of them of ‘*Ca. P. asteris*’ strains AY-WB and OY-M and of ‘*Ca. P. australiense*’ but only 4% of ‘*Ca. P. mali*’ genes are encoded in such structures. Even if this provisional assignment of genes to phages or PMUs will show overlaps, there is no doubt that both events took place in phytoplasma genomes and that it seems unlikely that PMUs are the major driving force for duplicated genes in ‘*Ca. P. mali*’ that encodes only one *tra5* transposase as central element of PMUs. Interestingly, no PMUs have been identified in *A. laidlawii*.

The impact of other genetic elements such as reverse transcriptases remains unclear. At least ‘*Ca. P. asteris*’ strain OY-M and ‘*Ca. P. australiense*’ encode a reverse transcriptase (ST1). The presence of such a gene usually indicates a mobile element such as a retrotransposon or retrovirus [61]. The

result of these integration and recombination events can be visualized by cumulative GC analysis that shows irregular skews for the ‘*Ca. P. asteris*’ strains and ‘*Ca. P. australiense*’ but only in parts for ‘*Ca. P. mali*’ and not for *A. laidlawii* (Figure 2).

All these events enlarge the condensed phytoplasma genomes [49], but the enlarged genome size does not necessarily expand the number of encoded functions.

Different levels of genome condensation [20] and the putative phage-derived horizontal transfers of genetic material [56] or transposon-driven duplication events [49] result in different genome sizes in phytoplasmas. Insights for the COG (cluster of orthologous groups) assignment [62] of the protein-coding content of the four phytoplasmas show (Figure 3) that the majority of the functional categories is affected by these driving mechanisms and gives the impression that they share only for few categories a similar content. The largest category without any COG assignment may indicate the individual evolution of each phytoplasma but also an amount of genes, which are resulting from over-prediction or by representing pseudogenes. The analysis of the shared genetic content provides insights into the general demands on the host while differences may reflect strain-specific capabilities.

### 3. Replication, Recombination, and Repair

**3.1. Replication Proteins and *oriC*.** All essential genes necessary to generate the bacterial replication fork are encoded in the four genomes. These comprise the proteins forming the prereplisome such as DnaA, which binds at the *oriC*, the interacting DnaB/DnaC complex, and DnaD [63]. The phytoplasmas also share the DNA gyrase (GyrAB) and the DnaB helicase performing the bidirectional DNA unwinding [64]. The putative modulator PmbA of the DNA gyrase can influence this process.

Other encoded proteins, such as DnaG, probably synthesise initial primers at the site of the prereplisome [65], which is needed for the polymerase III heteromultimer. The genes encoding the proteins forming the DNA polymerase III holoenzyme are also present. This protein set contains PolC (alpha subunit, leading strand), DnaE (alpha subunit, lagging strand), DnaN (beta subunit), DnaX (tau subunit), and HolAB (delta and delta prime subunit). The putative assignment of PolC to the leading strand and the DnaE subunit to the lagging strand appears to be likely with respect to the Gram-positive ancestors [15, 66].

All phytoplasma genomes encode a NAD-dependent DNA ligase (Lig) for sealing breaks in DNA that occur, for example, during the discontinuous synthesis of the lagging strand. An ATP-dependent DNA ligase has not been annotated, a feature shared with *A. laidlawii* (Acc. no. CP000896). Other proteins involved in replication such as PriA responsible for the restart of stalled replication forks or acting at the occurring RNA/DNA hybrids (RnhC, NrdFA) are also present.

Some of the genes including *dnaB*, *dnaG*, and others involved in replication such as the single-stranded binding protein (*ssb*), the dna primase (*dnaG*), and the delta subunit

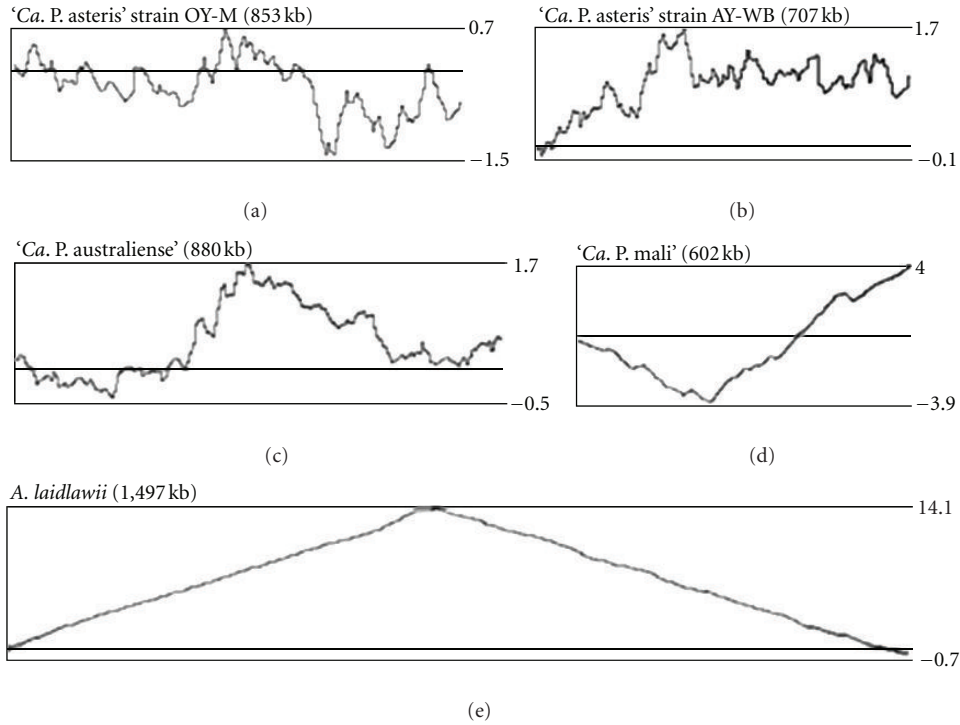


FIGURE 2: Cumulative GC skew analysis  $[(C - G)/(G + C)]$  of the chromosomes of the two ‘*Ca. P. asteris*’ strains OY-M and AY-WB, ‘*Ca. P. australiense*’, ‘*Ca. P. mali*’, and *A. laidlawii*. A window size of 5000 bases was used for calculation. Maxima and minima values obtained are indicated.

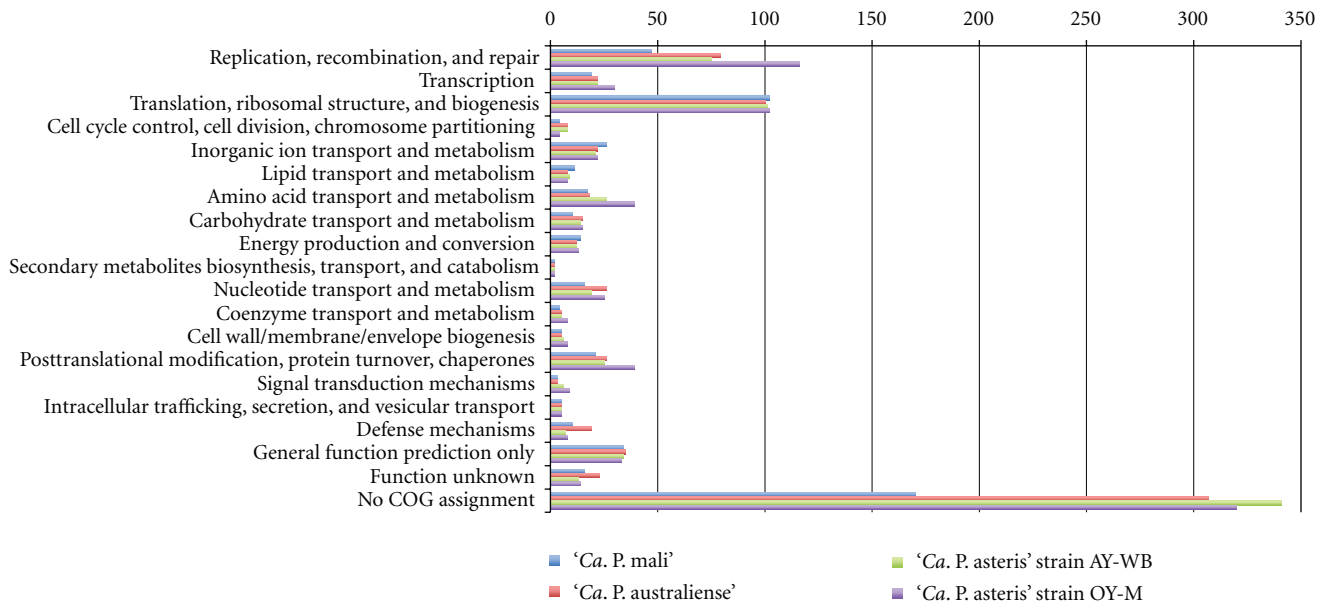


FIGURE 3: COG assignment of the deduced proteins of the four phytoplasma genomes. Single deduced proteins can show more than one assignment to the COGs database.

of the DNA polymerase III (*polC*) are present as multicopy genes. The origin of these multiple copies is still in discussion but they can be assigned to PMU or phage origin (see Section 2). In spite of this discussion it is notable that the *dnaB* gene is annotated at least twice in phytoplasmas. A situation that is also present in *A. laidlawii* (Acc. no. CP000896). The

annotated DnaB2 shows significant differences to DnaB1 in phytoplasmas. DnaB1 contains the DnaB-like helicase N-terminal domain (Pfam entry: PF00772.15) and the corresponding C-terminal domain (Pfam entry: PF03796.9) while *dnaB2* just partially matches the DnaB 2 family entry for replication initiation and membrane attachment

(Pfam entry: PF07261.5). A tandem arrangement of *dnaB2* and *dnaC* is encoded in all four genomes. Other tandem arrangements of single copy genes such as *dnaAN* or *gyrAB* and *nrdA/nrdF* are also present in phytoplasmas.

The coding for the suggested telomere resolvase is restricted to 'Ca. P. mali' [20]. Apart from that *orf*, the genetic environment of the genes involved in replication varies only in the copy number (ST2).

The *oriC* of the majority of bacterial chromosomes lies close to the overall minimum of the cumulative GC-skew [67, 68]. As phytoplasma chromosomes are heavily influenced by rearrangements and integration events that result in irregular cumulative skews and thus do not show a distinct minimum, an exception is the linear chromosome of 'Ca. P. mali' (Figure 2) [20, 49] that seems to be less influenced by such events and shows a shifted cumulative GC-skew with a minimum close to the sequence upstream the *dnaA* gene (Figure 2) [20]. The other three chromosomes show little pronounced minima close to the location of *dnaA*, indicating the presence of a putative *oriC* [20, 49]. This assignment is supported by several other approaches such as Oriloc [20, 49, 69] or OligoSkew analyses [20] and the typically decreased G + C content at the proposed *oriC*, which is present at the intergenic regions of the phytoplasmas flanking *dnaA*. They contain long AT-stretches of at least 30 bases (31, 33, and 48 bases in strain OY-M; 32 and 48 bases in strain AY-WB; 161 and 39 bases present upstream PAA\_0839 in 'Ca. P. australiense'; 30, 33, 43, 87, and 53 bases in 'Ca. P. mali'). AT-stretches were limited in 'Ca. P. australiense' and 'Ca. P. mali' to positions upstream *dnaA* due to the coding density.

The *oriC* regions show only a weakly conserved synteny (Figure 4), a situation similar to that of other *Mollicutes* [70]. A typical switch in the coding strand preference at the *oriC* is present within this region and corresponds to a bidirectional replication. However, 'Ca. P. mali' shows a coding strand preference all over the chromosome [20].

The DnaA protein of *Escherichia coli* initially binds a variant of the DnaA-box consensus 5'-TTAT(C/A)CA(C/A)A-3' at the *oriC* region [71]. DnaA-box motifs flank the *dnaA* gene in several *Mollicutes* [70]. The consensus of the DnaA-box motif of *Mollicutes* such as *M. pulmonis*, *Spiroplasma citri*, *M. capricolum*, and *M. genitalium* (5'-TT(A/T)TC(C/A)ACA-3') differs from the consensus of *E. coli* in the positions of the variable nucleotides (Table 2) [70]. The seven annotated DnaA-boxes of *A. laidlawii* (Acc. no. CP000896) result in a consensus motif (5'-TT(A/T)T(C/T)(C/A)ACA-3'), which largely matches the suggested motif for other mollicutes [70], but also shows that the fifth position is more variable (A/C/T). Such a consensus sequence of the DnaA-box motif was not identified at the *oriC*-region of the phytoplasma chromosomes. The suggested DnaA-box motif for *Mollicutes* occurs in the neighbourhood of *dnaA* of 'Ca. P. asteris' strain OY-M (position 4878–4886, within the disrupted *dotG* gene; complement 848'134–848'142, encoding also *ugpA* at this position) but not within the intergenic regions flanking the *dnaA* gene. It is absent within the flanking intergenic regions of *dnaA* in strain AY-WB and also in 'Ca. P. australiense', where the occurrence is limited to a region (position 1503–1511) encoding *dnaN*. A similar situation can be found in

TABLE 2: Consensus of the DnaA-box motives.

Species/group	DnaA-box	Reference
<i>Escherichia coli</i>	5'-TTAT(C/A)CA(C/A)A-3'	[71]
<i>Mollicutes</i>	5'-TT(T/A)TC(C/A)ACA-3'	[70]
<i>A. laidlawii</i>	5'-TT(T/A)T(C/T)(C/A)ACA-3'	(CP000896)
Candidate motif proposed for phytoplasmas	5'-TTAGGAACA-3'	(this study)

'Ca. P. mali'. The mollicute motif is present within the *dnaN* gene (position 252'493–252'501) and within the sequence encoding a putative ABC transporter subunit for methionine (position complement 248'550–248'558). Strain OY-M and 'Ca. P. australiense' share the motif 5'-TTTTCAACA-3' while 'Ca. P. mali' contains 5'-TTATCAACA-3'. However, it was not possible to identify the proposed mollicute motif in strain AY-WB, its occurrence within sequences encoding genes, and the rare occurrence within this region which raises the question if the consensus (5'-TT(T/A)TCAACA-3') covers the sequence of the DnaA-box motif. It seems to be likely that the DnaA-box motif shows a higher variability. Other motifs occur with higher frequency such as 5'-ATAGGAACA-3' (eight times), more interesting 5'-TTAGGAACA-3' (three times) for 'Ca. P. mali', and are located within the intergenic regions flanking *dnaA*. The latter is a promising candidate present in all four genomes within the *oriC* region with at least one copy (strain AY-WB) but contains an unusual GG-coding at positions four and five. It is present in the intergenic region of OY-M for four times, of AY-WB once, 'Ca. P. australiense' five times, and 'Ca. P. mali' three times. However, experimental evidence is needed to verify this.

**3.2. Proteins Involved in Chromosome Partitioning.** Segregation and partitioning of the chromosomal DNA start immediately after the initiation of the replication. However, the identified repertoire for this process is sparse, and it is likely that some proteins assigned as conserved hypothetical proteins will be involved. Topoisomerases will decatenate the two chromosome copies.

All four genomes also encode a regulator belonging to stage V sporulation protein family G (ST3). It is essential for sporulation and specific to stage V sporulation in *Bacillus megaterium* and *B. subtilis* [72]. However, the function in phytoplasma is unknown, and the protein may be involved in other processes.

In OY-M *smc*-genes were annotated that encode chromosome segregation of ATPase-like proteins. This assignment remains unclear, and the assignment to a glycoprotein appears to be more likely (see below).

**3.3. Recombination.** The complement of genes necessary for resolving Holliday junction and recombination differs in phytoplasmas. 'Ca. P. mali' shows the most complete set for *rec*-dependent repair system. It encodes *ruvAB* but the endonuclease *ruvC* was not identified. Instead, RecU may have this function. The RuvABC enzymatic complex binds

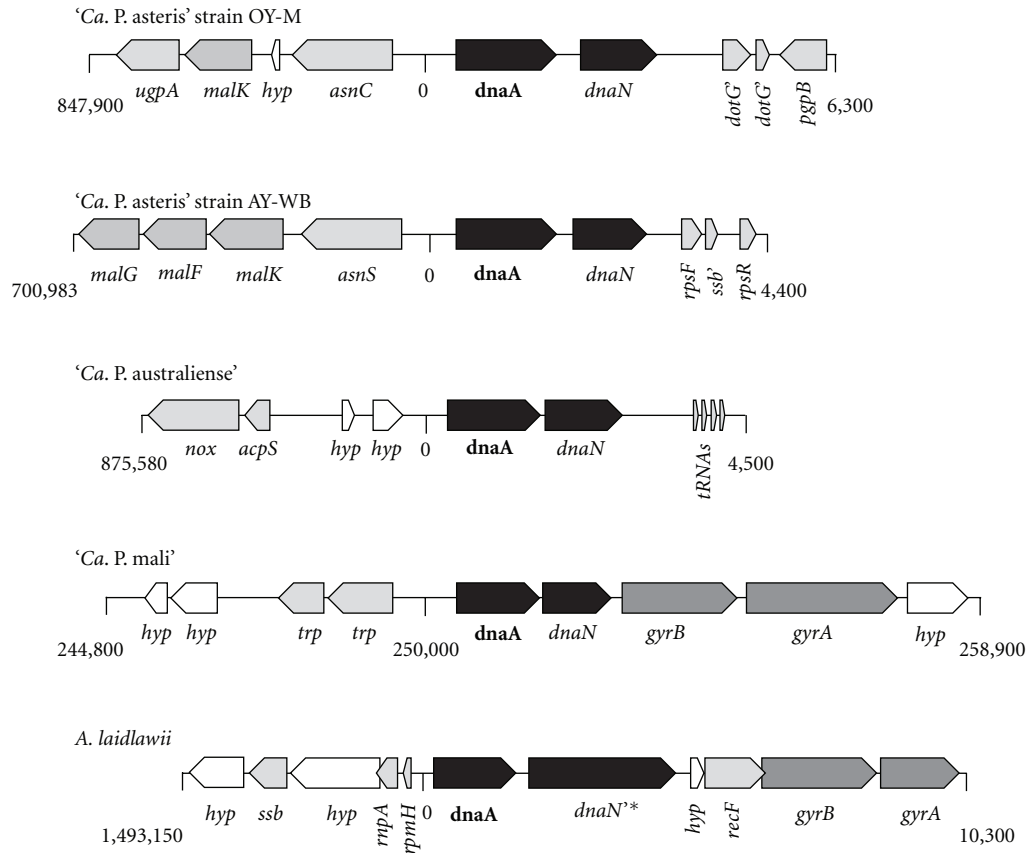


FIGURE 4: Genomic context at the suggested *oriC*. The *oriC* region is shown for the four phytoplasma chromosomes and *A. laidlawii* illustrating the weak conserved synteny within this region. Several genes are truncated (“ ”) and one gene probably destroyed by transposase integration (“\*\*”). Abbreviation of genes: *hyp*: hypothetical protein; *trp*: ABC transporter subunit. Genome positions were given for each region taking in account the circular organisation of four chromosomes.

to recombinational junctions catalysing strand cleavage and branch migration. This is needed for the repair of double-strand breaks, which result from the RecA-dependent pathway. A basic system encoded in ‘*Ca. P. mali*’ but not in the other three phytoplasmas consists of recombinase A, the ATP-dependent zinc-containing RecG-like helicase, and the recombinational DNA repair proteins O and R in contrast to the other three phytoplasmas. However, genes *recA*, *recU*, *ruvA*, and *ruvB*, which are absent in three phytoplasmas, do not belong to the essential gene set of *M. genitalium* but may have long-term influence on the chromosome maintenance [73].

The function of the RmuC-family protein containing an endonuclease domain remains unclear in this context [74]. It is annotated in phytoplasmas except for ‘*Ca. P. mali*’ (ST4).

Due to dispersed locations within the genome of ‘*Ca. P. mali*’ it seems to be unlikely that the *ruv*- and *rec*-genes derived from a horizontal transfer. One may speculate that if this gene cluster is still functional as a unit. In comparison to *A. laidlawii*, similarities such as the lack of *ruvC* are obvious but *refN*, *recF*, and *recD* are encoded.

Functional recombination machinery provides an advantage, if selection pressure induced by attacks from the environment or competition in the same ecological niche exists.

This aspect is addressed in the statement of “living with instability” of phytoplasmas [49], so differences in coding capacity would enable response to such selection pressure.

However, plasmid-derived proteins involved in replication such as Rep are encoded in asteris yellows phytoplasma chromosomes (PAM\_667, AYWB\_403).

**3.4. DNA Repair and Degradation.** UvrABCD proteins represent the genetic components of the nucleotide excision repair DNA repair system (ST5), which also is involved in SOS response. UvrA protein forms homodimers in the presence of ATP, binds to the DNA, and associates with UvrB. The two other units are formed by the nuclease UvrC and the helicase UvrD. They act jointly with the DNA ligase and complete the excision-repair process. The complete genetic repertoire is encoded in three phytoplasma genomes and *A. laidlawii* but is absent in ‘*Ca. P. mali*’ with exception of UvrD, which is probably also involved in other reactions due to its helicase activity.

All four phytoplasmas encode parts of the base excision repair pathway to repair certain types of premutagenic lesion (ST5). The genes *mutM* and *mutT* are present in all four phytoplasma genomes but additional genes of the *mut*-system such as the ATPase *mutLS* are missing. In contrast,



TABLE 3: Localization of the rRNA operons.

Product	OY-M	AY-WB	' <i>Ca. P. australiense</i> '	' <i>Ca. P. mali</i> '
16S rRNA	279401–280921	271740–273260	Complement (682149–683667)	264151–265657
23S rRNA	281174–284036	273513–276376	Complement (679085–681932)	265879–268741
5S rRNA	284074–284181	276414–276521 <sup>2</sup>	Complement (678945–679050)	268777–268888
16S rRNA	Complement (555991–557511)	Complement (496364–497884) <sup>1</sup>	Complement (863584–865102)	450392–451898
23S rRNA	Complement (552876–555738)	Complement (493248–496111) <sup>1</sup>	Complement (860519–863366)	452120–454982
5S rRNA	Complement (552731–552838)	Complement (493103–493210) <sup>2</sup>	Complement (860380–860484)	455018–455129

rRNA: ribosomal RNA.

<sup>1</sup>Gene was assigned to the opposite strand in contrast to CP000061.

<sup>2</sup>5S-rRNA gene was not annotated in CP000061.

they are present in *A. laidlawii*, which also carries *mutT* but lacks *mutM*. It is possible that other genes compensate the function of MutLS and enable the hydrolase and the glycolyse/lyase to function in combination with the other encoded proteins required. The annotated uracil-DNA glycolase would support the presence of such a basic system.

The genetic repertoire for the nucleotide excision repair DNA for damaged oligonucleotides is encoded in '*Ca. P. mali*', while misincorporated bases cannot be corrected by the nucleotide excision repair system. This will result in uncorrected bases, which include the deamination leading to a C to A transversion and may affect the G + C content of '*Ca. P. mali*'.

An incomplete restriction-modification system is encoded in strain OY-M but is absent in the closely related strain AY-WB and '*Ca. P. australiense*' and '*Ca. P. mali*'. This indicates that this system does not belong to the shared core set. It may thus be derived from a horizontal transfer.

**3.5. DNA Modification and Structure.** All phytoplasmas contain the bacterial core proteins encoding the type I topoisomerase TopA and the two gyrase subunits forming the type II topoisomerase (ST6). This protein content enables the phytoplasmas to relax supercoiled DNA by introducing single-stranded or double-stranded DNA breaks.

The high copy number of methyltransferases and of the nucleotide-binding proteins is remarkable. The methylation protects the DNA, for example, from restriction by endonucleases, a strategy, which is used to protect the integrity of the chromosomal DNA and to decrease the number of integration events. The high number of nucleotide-binding proteins corresponds to the latter scenario. This is in agreement with the high number of copies of the integration host factor HimA, which is required in site-specific phage recombination [75] and is part of the complex transposons such as PMU1 [59]. The copy numbers of these genes differ even between closely related strains. Only a single copy of *himA* is present in '*Ca. P. mali*', corresponding to the putative single PMU [60].

## 4. Transcription

**4.1. RNA Polymerase and Sigma Factors.** RNA polymerase subunits for alpha, beta, beta', and omega chain are encoded in all 4 genomes (ST7). The sigma subunit is present with

several paralogs, which is not unusual for bacterial genomes. The regulators RseABC and the sigma factors related to heat shock (RpoH), metabolic functions (RpoN), or stress (RpoS) were not identified. It is remarkable that the genomes of strain OY-M and of '*Ca. P. australiense*' encode a high number of *fliA* genes whereas '*Ca. P. mali*' encodes only a single gene. The small sigma factor FliA is known to be involved in transcriptions of operons involved in chemotaxis, motility and biofilm formation [76, 77]. However, there are no hints available on these abilities except for transcription in phytoplasmas.

**4.2. Factors Affecting the RNA Polymerase.** A common set of genes encoding proteins for transcription regulation and elongation is shared (ST8). It consists of the transcription elongation factor GreA and the N utilization substance protein involved in prevention and enhancement of transcription termination [78]. The transcription coupling factor Mfd [79] and the transcriptional termination factor Rho [80], like in *Mycoplasma* species, have not been identified in phytoplasmas.

**4.3. Transcriptional Regulators.** Only a sparse repertoire of transcriptional regulators is present in phytoplasmas. The negative regulator of class I heat shock proteins HrcA and putative cold shock proteins are encoded in all four phytoplasma genomes (ST9). Regulators of two-component systems were not annotated in these genomes.

## 5. Translation

**5.1. rRNA-Operons.** Two rRNA operons are encoded in the complete determined phytoplasma genomes. This applies for all phytoplasmas [81] and *A. laidlawii* (Acc. no. CP000896). The positions of the rRNA genes (Table 3) were recalculated using the actual version of the RNAmmer (v.1.2) software [82]. However, it should be taken in account that the prediction by RNAmmer, which is based on HMM and BLAST, of start and stop position can differ by about 10 nt. A switch in strand preference of the rRNA operons with respect to the proposed *oriC* and *terC* is only present in the two aster yellows strains.

**5.2. tRNA Synthetases.** Aminoacyl tRNA synthetases catalyse the esterification of a specific amino acid or its precursor

to one of all its compatible cognate tRNAs to form an aminoacyl-tRNA. They are essential for the translation process.

In contrast to the other tRNA synthetases (ST10), the aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunits ABC (*gatABC*) are absent in all phytoplasma genomes and *A. laidlawii*. This absence separates the four phytoplasma genomes from other *Mycoplasma*, *Ureaplasma*, and *Mesoplasma* species.

**5.3. Transfer RNAs.** Transfer RNAs are essential for the transfer of a specific active amino acid to a growing polypeptide chain during translation. The complete set of tRNAs was recalculated for the four genomes by tRNAscan-SE v.1.23 using the bacterial model [83]. The qualitative endowment of tRNAs is shared by the phytoplasmas but there are quantitative differences. Most different is '*Ca. P. australiense*' by the presence of an additional gene for tRNA-Gln/His/Leu and only a single copy for tRNA-Val (ST11). Transfer RNA coding for the usage of selenocysteine was not identified in the four genomes and *A. laidlawii*.

**5.4. tRNA/rRNA Modification Factors.** All phytoplasmas share a similar set of modification factors (ST12), except for RimM, which was identified only in '*Ca. P. mali*'. It is also present in *A. laidlawii*. RimM is essential for the efficient processing of 16S rRNA in *E. coli* [84]. It has affinity for free ribosomal 30S subunits but not for 30S subunits in the 70S ribosomes. The absence in the other three phytoplasmas and *Mycoplasma* species indicates the distinct phylogenetic position of '*Ca. P. mali*' but also the higher relatedness to the achleoplasmas than the other three phytoplasmas.

**5.5. Translation Factors.** All four genomes share a common set of translation factors including LepA (elongation factor 4) (ST13). LepA is a ribosomal back translocase related to EF-G and EF-Tu and is supposed to recognize ribosomes after a defective translocation reaction. This reaction allows EF-G to translocate the tRNAs correctly [85].

PrfA and PrfB mediate translation termination. The presence of these proteins is in accordance with the observation that not only the four fully sequenced phytoplasma genomes but all phytoplasmas use the bacterial code. PrfB is absent in *Mycoplasma* species, which translate UGA as tryptophan rather than using it as a stop codon [86].

**5.6. Ribosomal Proteins and Modifying Factors.** The environment for genes encoding ribosomal proteins consisting of 32 large subunits and 20 small subunits was identified in all phytoplasma genomes (ST14), in addition to some modifying enzymes (ST15). The genes encoding the large subunit protein RplY (L25) and the small subunit protein RpsA (S1) were not identified. However, RpsA was identified in *A. laidlawii*.

Some genes seem to be absent in phytoplasmas such as the ribosomal protein L11 methyltransferase (PrmA) or FtsJ. They also are absent in other mollicutes. The glutathione synthetase RimK present in *M. pneumoniae* and *M. genitalium* is also absent in all phytoplasmas. 50S ribosomal

stability factor encoded by the subunits *engABCD* is present in all phytoplasmas.

**5.7. mRNA Degradation.** All four genomes encode proteins similar to YkqC from *Bacillus licheniformis* (YP\_078844.2) encoding the essential RNase J1/J2 involved in mRNA degradation (ST16).

**5.8. Heat Shock Proteins and Chaperons.** The four phytoplasmas share a basic gene set containing the Hsp70-type (*dnaK*, *grpE*), the Hsp60-type (*groEL*, *groES*) chaperone system [87], the chaperone *dnaJ*, and the trigger factor (ST17). The small heat shock protein IbpA (Hsp20) that binds to aggregated proteins and is present in other phytoplasmas [88] was not identified in '*Ca. P. mali*'.

Other encoded cytosolic proteases such as Lon-protease that hydrolyses ATP and unfolds bound substrates are encoded in all four phytoplasmas. These proteins are involved in various cellular activities [89].

## 6. Cell Envelope and Cell Division

**6.1. Proteins Involved in N-Acetylglucosamine and Murein Biosynthesis.** Proteins involved in N-acetylglucosamine and murein biosynthesis have not been identified in phytoplasma genomes. The absence of these genes is in accordance with the lack of a cell wall. However, all four genomes encode MraW (ST19), which is a S-adenosyl-dependent methyltransferase also described to show activity toward substrates associated with membrane components [90].

**6.2. Proteins Involved in Lipopolysaccharide Synthesis.** Only *rfaG* encoding for core functionality in lipopolysaccharide (LPS) synthesis was identified in '*Ca. P. asteris*' strain OY-M and '*Ca. P. mali*'. RgaG acts as a UDP-glucose: (heptosyl) LPS  $\alpha$ -1,3-glucosyltransferase in *E. coli* [91] releasing UDP and H<sup>+</sup>. This enzyme is involved in lipid A-core biosynthesis (ST20). Other proteins of this pathway were not identified.

**6.3. Cell Division Proteins and Regulators.** In enterobacter, cell division is mainly mediated by *fts* and *min* genes that determine cell shape. In phytoplasmas cell division proteins FtsY and FtsH were identified. Two additional genes associated with cell division were detected (ST21). The impact of the glucose-inhibited division (Gid) proteins encoded by phytoplasmas on the regulation of the cell division processes remains unclear. GidA is reported to be involved in 5-carboxymethylaminomethyl modification of the wobble uridine base in some tRNAs [92, 93] whereas GidB, which is missing in '*Ca. P. mali*', is suggested to act as S-adenosyl-L-methionine- (SAM-) dependent methyltransferase [94]. Both genes are encoded in the majority of bacterial genomes and also encoded in the genome of *A. laidlawii*. However, it was shown by mutagenesis of *M. genitalium* and *M. pneumoniae* that GidB is not essential [95]. It was suggested that the GidB of *E. coli* is specific for sterol and/or lipid substrates, but it also seems to be possible that GidB is specific for nucleic acids [96].

## 7. Membrane Proteins, Secretion, and Transport

**7.1. Porins and Outer Membrane Proteins.** Phytoplasma membrane proteins are in direct contact with their environment due to the lack of a cell wall. Membrane proteins such as Vmp1 show a wide diversity in accordance with their importance for phytoplasmas [97]. A subset of the membrane proteins are the abundant immunodominant proteins, which also show a high diversity [98]. Proteins such as the antigenic membrane protein AmP interact with the insect cell microfilaments and contribute to the insect vector specificity [99]. They are under positive selection pressure [100]. AmP is encoded in the genomes of the two ‘*Ca. P. asteris*’ strains. A putative homolog is also present in ‘*Ca. P. australiense*’, while ‘*Ca. P. mali*’ encodes the immunodominant protein ImP (ST22) [101]. However, *imp* and *amp* were also identified in ‘*Ca. P. asteris*’ strain OY-W [102].

These proteins represent only a small part of the predicted membrane proteins that are characterized by at least one transmembrane (TM) helix (ST23). Integral membrane proteins and secreted proteins have been efficiently predicted by Phobius [103] from the annotated proteins in all four genomes. The ‘*Ca. P. asteris*’ strains OY-M and AY-WB carry 184 and 169 proteins containing at least one TM domain but no signal peptide (SP). The large chromosome of ‘*Ca. P. australiense*’ encodes 181 of such proteins and ‘*Ca. P. mali*’ with the shortest chromosome 145. Some membrane proteins carry a SP in addition that indicates a cleavage during protein translocation [98]. A low number of these proteins have been predicted in phytoplasmas: 12 in ‘*Ca. P. asteris*’ strain OY-M, 9 in strain AY-WB, 10 in ‘*Ca. P. australiense*’, and 5 in ‘*Ca. P. mali*’ (ST24).

The majority of membrane-associated proteins involved in transport and metabolism will be treated within the following sections.

**7.2. Sec-Dependent Pathway (General Signal-Dependent Export Pathway).** The *sec*-dependent pathway represents the best-characterized secretion system of phytoplasmas. The four phytoplasma genomes share a common gene set of *sec*-genes. (ST25). Genes encoding proteins of the general secretion pathways (T2SS) were not identified and also the twin arginine pathway for the secretion of folded proteins and their cofactors is absent. This can be explained by the absence of a second outer membrane. Other secretion pathways of phytopathogenic bacteria such as the type III secretion system (T3SS), T4SS, or pili have not been identified [104]. There is weak indication for a deduced IcmE-like membrane protein involved in a T4SS (*orf552* in ‘*Ca. P. australiense*’) [50] but further analysis is needed.

Secreted phytoplasma proteins may directly interact with their hosts and may in this way manipulate or weaken the plant host or insect vector without a needled system as provided by T3SS. Examples for the phytoplasmas are the secreted proteins *tengu* and SAP11 of ‘*Ca. P. asteris*’ strains OY-M and AY-WB, respectively [105]. For *tengu* it has been shown that the protein expression in *Arabidopsis thaliana* results in dwarfism [106]. The *sec*-dependent pathway is

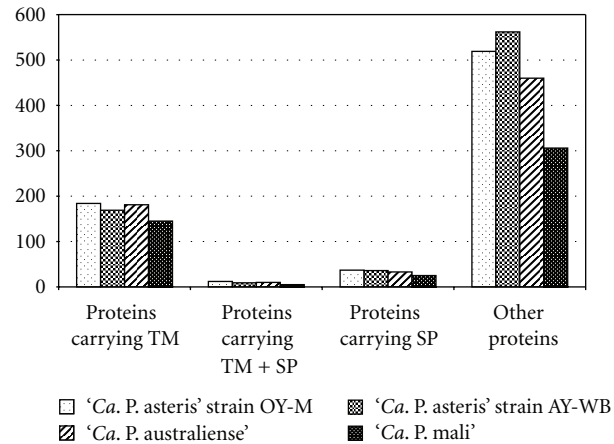


FIGURE 5: Phobius predictions of the four phytoplasma genomes. Abbreviations: TM: transmembrane helices; SP: signal peptide.

encoded in all four genomes by *secYEG* forming the integral membrane pore complex (Figure 5) and completed by *secA* and *yidC* [107]. YidC mediates the membrane insertion/assembly of the inner membrane proteins [108].

Genes encoding SecB, SecD/F, SecG, and YajC were not identified in the four genomes (ST25). GroEL may fulfil the function of the chaperone SecB. The effect of the absence of bifunctional SecD/F protein remains unclear due to the lack of information on their exact function. They might act to facilitate the protein export [109, 110]. The gene encoding the integral membrane protein *secG* is present in *A. laidlawii* but absent in phytoplasmas. A similar situation is present for YajC, which is supposed to stabilize the insertion of the SecA-preprotein complex in other bacteria.

The signal peptide mechanism seems to be similar to *E. coli*, and prediction software such as SignalP [111] or PSORT developed for Gram-negative bacteria [112] can be applied for the prediction of proteins secreted by the *sec*-dependent pathway [107] and other software such as Phobius [103] limiting the analysis on proteins carrying a signal peptide but no transmembrane regions (Figure 5). The ‘*Ca. P. asteris*’ strains OY-M and AY-WB carry 37 and 36 predicted secreted proteins containing a SP, while ‘*Ca. P. australiense*’ encodes 33 and ‘*Ca. P. mali*’ 25 (ST26).

**7.3. Signal Recognition Particle (SRP) Pathway.** The SRP is a ribonucleoprotein, which was identified in all organisms. It acts in targeting translating ribosomes to the SecYEG translocation complex for cotranslational protein translocation over the membrane [113]. The SRP complex consists of the RNA component (4.5S RNA, *ffs* gene product) and the protein component (the *ffh* gene product). This complex is required for targeting some integral membrane proteins to the membrane for cotranslational integration [114]. Additional proteins are required for the functional translocation system. They comprise YidC (see above), which is bound to the SecYEG translocation complex and FtsY. The latter serves as docking protein for the SRP-ribosome complex synthesising the nascent peptide [115].

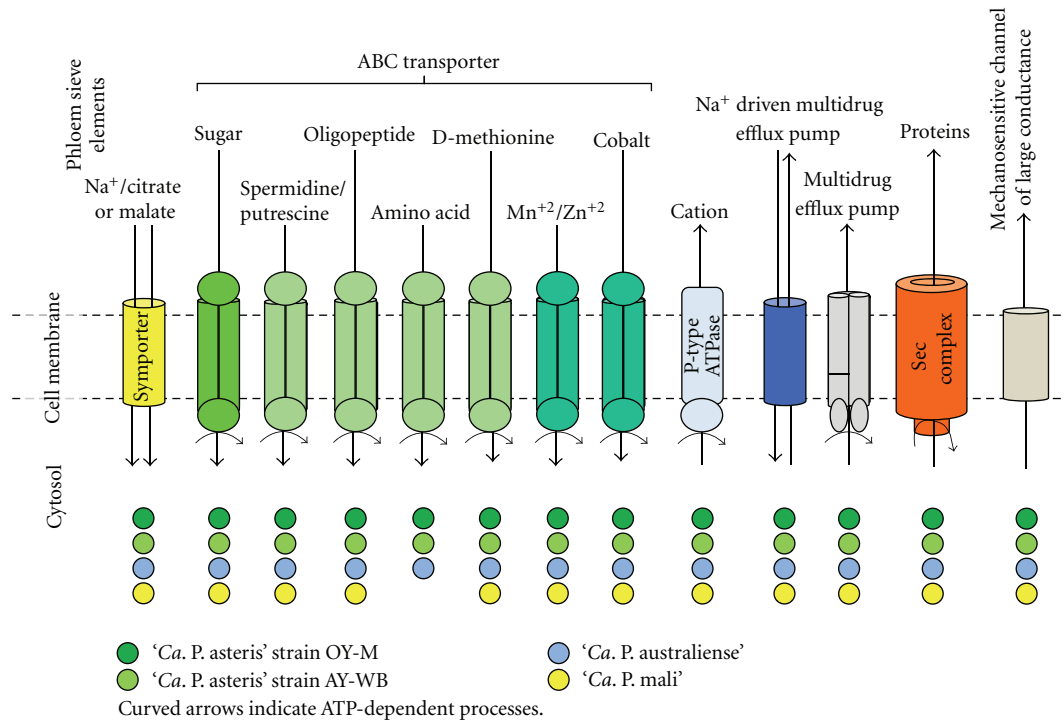


FIGURE 6: Symporter, transporter, ion translocation, and secretion.

All units forming this protein complex were encoded in the four phytoplasmas (ST27) except the *secG* and the *ffs* gene for 4.5S RNA, which were also not identified in achleplasmas. However, it is difficult to identify 4.5S RNA that is encoded by eubacteria, archbacteria, and *Mycoplasma* species [116].

**7.4. ABC Transporters and Symporters.** All four phytoplasmas depend on the uptake of essential compounds from their hosts (Figure 6) [48]. They encode a common core set of ABC transporters (ST28). This includes the clustered spermidine/putrescine transport system (*potABCD*), a Mn/Zn transport system (*znuACB*), sugar transport system (assigned as *malEGFK*), a dipeptide/oligopeptide transport system (assigned as *dppCBADF*), and a methionine transport system (*metNQI*). An additional amino acid transport system (*artQPIM*) is encoded in OY-M (probably duplicated), AY-WB, and 'Ca. P. australiense' but is absent in 'Ca. P. mali'. The genes for *znuACB* are present in two copies in 'Ca. P. mali' due to the location in the terminal inverted repeats.

The template of the sugar transport system remains unclear. The uptake of maltose, trehalose, sucrose, and palatinose was suggested, and sucrose appears to be most likely [50, 117] with respect to the presence in the phloem and the corresponding protein machinery. *Spiroplasma citri*, another phloem colonizing parasite, is able to utilize glucose, fructose, and trehalose [118].

The subunits in phytoplasmas show similarities to maltose/maltodextrin transporters but also to glycerol-3-phosphate transporters (*ugpBEAC*). Both protein sets are highly homologous and might be functionally exchangeable,

but show differences with respect to the growth substrates [119]. The import of glycerol-3-phosphate would also offer a phosphate source for the phytoplasmas.

The phytoplasma dipeptide/oligopeptide transport system shows sequence differences in the substrate-binding unit that may indicate differences in substrate specificity.

In addition, several incomplete ABC transporters (ST29) are present in phytoplasmas. Subunits of the above mentioned transporters of dipeptide/oligopeptide transporters, amino acids, and cobalt are present in the genome. Incomplete transporter units might be completed by subunits of other peptide and amino acid transporters. The functionality of cobalt import system remains unclear because the periplasmic subunit was not identified. However, the permease (*cbiQ*) and ATPase subunits (*cbiO*) are present in all phytoplasma genomes.

A putative thiamine transporter subunit is encoded in all four phytoplasma genomes (ST30). The corresponding peptide sequences were assigned by InterPro [120] to the YuaJ-family (IPR012651). Many YuaJ-family members have been assigned as ATP-independent thiamine transporter, which are involved in regulation of the thiamine pyrophosphate (TPP) concentration [121]. The import is probably proton coupled. Thiamine is phosphorylated at the membrane by thiamine kinase (ThiK) in *Bacillus cereus* [122]. ThiK has not yet been identified in mollicute genomes. This raises the question if thiamine is imported or if it is an already phosphorylated substrate. The thiamine monophosphate kinase ThiL (syn. ThiJ) is encoded in the four phytoplasma genomes (PAM\_137, AYWB\_584, PAa\_0495, ATP\_00237). This finding supports the idea that a phosphorylated thiamine substrate

or an unassigned protein has to fulfil the function of ThiK in phytoplasmas.

Beside the ABC transporters, malate or citrate/Na<sup>+</sup> symporters are encoded in all four genomes that may provide an important carbon source [49] (ST31).

Several other membrane-located proteins are involved in transport in all four genomes and belong to the core gene repertoire of phytoplasmas (ST32). They contain a multidrug efflux pump such as MdlAB and the Na<sup>+</sup>-driven multidrug efflux pump NorM. The genes assigned as *norM* differ in sequence similarity. The NorM protein from '*Ca. P. mali*' has only one ortholog in '*Ca. P. australiense*' (PAa.0171).

Another common feature is the P-type ATPases exporting cations such as magnesium, calcium, and cadmium (ST33). Notably, mutants deficient in the single P-type ATPase encoded in *Spiroplasma citri* are affected in their growth capacity [123].

Finally, the regulation of the osmotic pressure by opening the membrane lipid bilayer to prevent cell disruption and death [124] is given by the mechanosensitive channel formed by the MscL protein in phytoplasmas (ST34).

## 8. Metabolism

**8.1. Carbohydrate Metabolism Glycolysis.** Embden-Meyerhof-Parnas pathway was suggested to be the major energy-yielding pathway in phytoplasmas [48] despite the apparent lack of hexokinase (glucose phosphorylating) and a sugar-specific phosphotransferase system (PTS) mediating a phosphorylated hexose to enter glycolysis. One promising candidate is the glucosyltransferase GtfA, which was predicted in '*Ca. P. australiense*' first. The assignment of the deduced protein sequence and the ortholog of OY-M were confirmed by InterPro (IPR022527 sucrose phosphorylase, GftA). The GtfA, which is probably better described as disaccharide glucosyltransferase or sucrose phosphorylase, allows the formation of  $\alpha$ -D-glucose-1-phosphate from phosphate and sucrose [125], which is often a predominant sugar in the phloem.  $\alpha$ -D-glucose-1-phosphate is the entry compound of glycolysis. GtfA may compensate the absence of a hexokinase and PTS system. However, GtfA or a similar phosphorylase are not a general trait of phytoplasmas, since they are absent in the genomes of closely related AY-WB and in '*Ca. P. mali*'. Thus, the observed differences in glycolysis among phytoplasmas may arise from genome plasticity as suggested by the close proximity of prophage-related elements and *gtfA* in '*Ca. P. australiense*'.

Theoretically, the uptake of phosphorylated hexoses would overcome this problem. Candidates for such sugar phosphates are trehalose-6-phosphate, sucrose-6-phosphate, and  $\beta$ -D-fructose-6-phosphate. The transporter complements of the phytoplasmas may contain uptake systems for importing these phosphorylated di- and monosaccharides from the environment. For example trehalose-6-phosphate has to be monomerized prior to entry into the glycolysis. The breakdown would occur within the phytoplasmas mediated by hydrolases/phosphatases. At least one phosphatase each of subfamily IIIa and of subfamily IIb is encoded within the four phytoplasma genomes (ST35). These phosphatases are

poorly characterized so far. A single copy of the IIIA subfamily of the haloacid dehalogenase (HAD) superfamily of hydrolases representing hypothetical proteins is encoded in each phytoplasma genome. Most characterised members of this subfamily and of the HAD superfamily are phosphatases. This protein family consists of sequences from fungi, plants, cyanobacteria, Gram-positive bacteria, and *Deinococcus* (according to IPR010021 entry).

Functional interpretation of the second hydrolase of the Had superfamily hydrolase subfamily IIb may be more straightforward. Members of the Had superfamily hydrolase subfamily IIb are encoded by at least one gene ('*Ca. P. mali*') in each genome. They encompass trehalose-6-phosphatase, plant and cyanobacterial sucrose phosphatase, and a closely related group of bacterial and archaeal orthologs, eukaryotic phosphomannomutase (according to IPR006379 entry). If these proteins function as trehalose-6-phosphatase, phytoplasmas could use  $\alpha$ ,  $\alpha$ -trehalose-6-phosphate, and phosphate to produce glucose-6-phosphate and  $\beta$ -D-glucose-1-phosphate, which are the entry molecules of glycolysis.

If sucrose-6-phosphate is used as a substrate, sucrose-6-phosphatase may generate glucose-6-phosphate and fructose [126]. The utilization of trehalose-6-phosphate and/or sucrose-6-phosphate appears to be likely due to the phosphoglucose isomerase encoded in all four genomes. This step would be unnecessary, if fructose-6-phosphate is available. However, it should be considered that only trace amounts of trehalose-6-phosphate are present in higher plants on average. The impact of a phytoplasma infection on plant metabolism cannot be estimated so far, but trehalose-6-phosphate is a signaling molecule in plants with strong regulatory effects on metabolism, growth, and development [127–129]. The sucrose-6-phosphate concentration in the phloem is unclear. It is an interesting scenario that sucrose and trehalose compounds could be utilized depending on their availability in phloem and hemolymph.

The general upper part of the glycolysis (energy demanding) is encoded within all four phytoplasma genomes [20, 48–50] starting with  $\alpha$ -D-glucose-6-phosphate converted to  $\beta$ -D-fructose-6-phosphate by phosphoglucose isomerase (Pgi) and ATP-dependent formation of two molecules  $\beta$ -D-fructose-1,6-bisphosphate by phosphofructose kinase (PfkA). Subsequently, fructose-bisphosphate aldolase (Fba) catalyses the formation of D-glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate. The latter is suggested to enter glycerophospholipid metabolism, while D-glyceraldehyde-3-phosphate is channelled into the energy yielding part of the glycolysis (Figure 7). The interconversion of these C<sub>3</sub>-intermediates is performed by triosephosphate isomerase (TpiA).

Dihydroxyacetone phosphate can also be generated by a conserved kinase related to dihydroxyacetone kinase (DhaK). However, it remains unclear if this kinase can also act in the opposite direction as a transferase.

Except for '*Ca. P. mali*' [20] the protein components of the lower part of glycolysis (energy yielding) from D-glyceraldehyde-3-phosphate to pyruvate are encoded in the analyzed phytoplasma genomes (ST35): glyceraldehyde-phosphate dehydrogenase, phosphoglycerate kinase, and

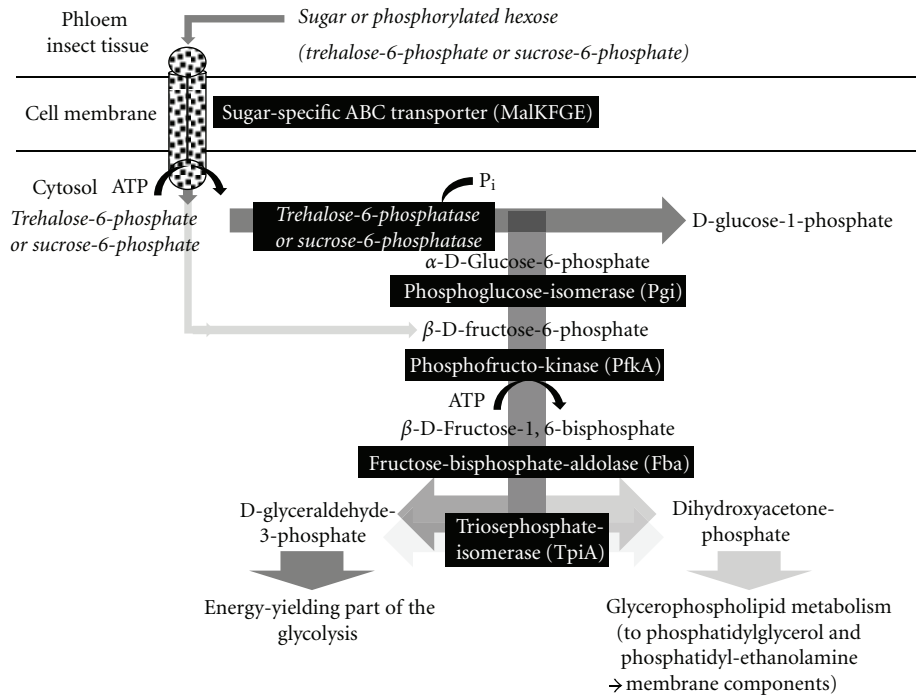
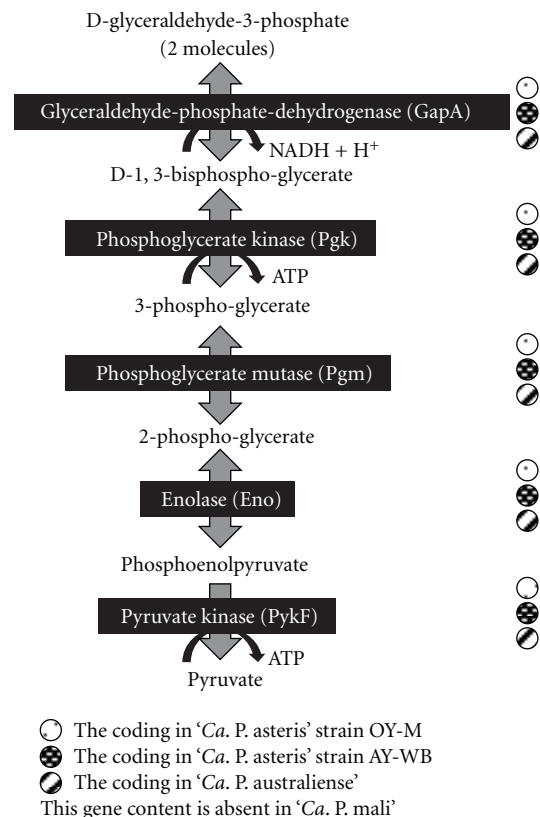


FIGURE 7: The sugar uptake and upper part of the glycolysis (energy investment) including trehalose-6-phosphate and  $\beta$ -D-fructose-6-phosphate as suggested entry substrates (steps indicated in italics).

mutase, enolase, and pyruvate kinase (Figure 8). Notably, gluconeogenic phosphoenolpyruvate synthase (PpsA) is lacking in all four phytoplasmas (Figure 8). Other strategies to obtain host-derived ATP such as ATP/ADP translocase known from *Chlamydia* species [20, 48, 130] and also the arginine dihydrolase pathway [131] have not been identified in phytoplasmas.

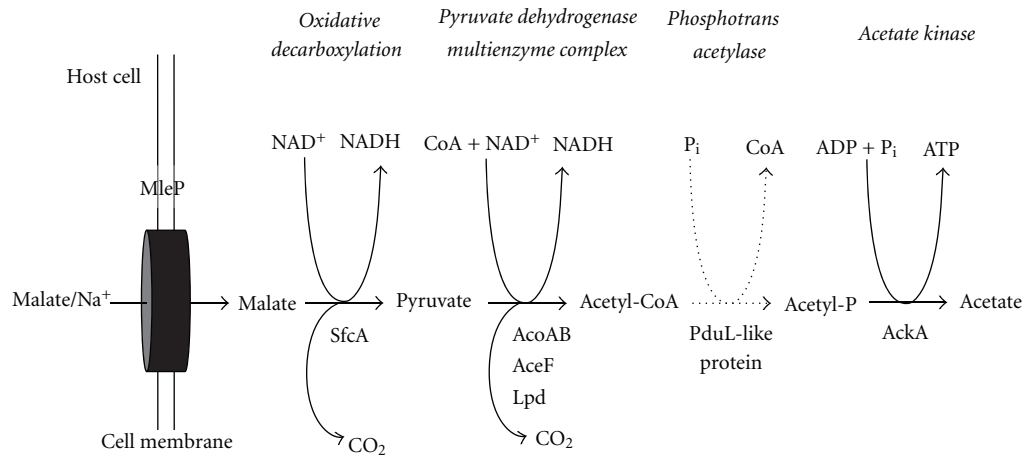
8.2. An Alternative Energy-Yielding Pathway Deduced from the Genome Sequences. Carbohydrate metabolism is one of the most important physiological traits of the phytoplasmas. Alternative pathways have to be considered that could compensate for the lack of a complete glycolysis in ‘*Ca. P. mali*’ (Figure 9, ST36). A pathway for malate conversion to acetate is potentially encoded in all four genomes. Uptake of malate is enabled by the symporter MleP [20, 48]. Malate can be oxidatively decarboxylated to pyruvate by the malic enzyme ScfA. Pyruvate would then also be oxidatively decarboxylated to acetyl-CoA by the pyruvate dehydrogenase multienzyme complex.

In the case of ‘*Ca. P. mali*’ pyruvate might be generated by an additional way. Here, an aldolase (Eda) is predicted, serving two possible functions. A 4-hydroxy-2-oxoglutarate aldolase (EC: 4.1.3.16) (KHG-aldolase) would catalyze the interconversion of 4-hydroxy-2-oxoglutarate into pyruvate and glyoxylate. Phospho-2-dehydro-3-deoxygluconate aldolase (EC: 4.1.2.14) (KDPG-aldolase) would catalyze the interconversion of 6-phospho-2-dehydro-3-deoxy-D-gluconate into pyruvate and glyceraldehyde-3-phosphate. In both cases, pyruvate would be formed independently from glycolysis.



○ The coding in ‘*Ca. P. asteris*’ strain OY-M  
 ● The coding in ‘*Ca. P. asteris*’ strain AY-WB  
 ◐ The coding in ‘*Ca. P. australiense*’  
 This gene content is absent in ‘*Ca. P. mali*’

FIGURE 8: The energy-yielding part of the glycolysis encoded in the asteris strains and ‘*Ca. P. australiense*’ but absent in ‘*Ca. P. mali*’.



Protein names and products:

MleP	Malate/Na <sup>+</sup> symporter
SfcA	Phosphate acetyltransferase
AcoAB(syn. PdhAB)	Pyruvate dehydrogenase E1 component, $\alpha$ and $\beta$ subunit
AceF(syn. PdhC)	Pyruvate dehydrogenase E2 comp. (dihydrolipoamide acetyltransferase)
Lpd(syn. PdhD)	Dihydrolipoamide dehydrogenase
PduL-like protein	Phytoplasmas encode a phosphotransacetylase similar to PduL but lack the common phosphotransacetylase Pta
AckA	Acetate kinase

FIGURE 9: Proposed energy-yielding pathway from malate to acetate.

It is also notable that the gene set necessary for the generation of coenzyme A (CoA) is only partially annotated in phytoplasmas. In *E. coli*, CoA is synthesized from (R)-pantothenate. This requires involvement of 4'-phosphopantothenoylcysteine decarboxylase and phosphopantothenoylcysteine synthetase (fused in *E. coli*, CoaBC), the phosphopantetheine adenylyltransferase (CoaD), and the dephosphocoenzyme A kinase (CoaE). CoaBC was identified in some *Mollicutes*, while CoaD is annotated in the majority of phytoplasmas and *A. laidlawii*. While genes encoding CoaBC and CoaD were not identified in phytoplasmas, CoaE (annotated as formamidopyrimidine-DNA glycosylase) is encoded in all four genomes. The intermediate dephospho-CoA is the substrate of CoaE. This finding supports the possibility that at least the last steps in CoA-biosynthesis are performed by the phytoplasmas, but it remains unclear whether alternative reactions for the other steps in the pathway are encoded in the genomes.

Hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) could be formed by HcaD, oxidizing NADH. H<sub>2</sub>O<sub>2</sub> may also be generated by the encoded superoxide dismutase and represent a potential virulence factor, as known from *M. pneumoniae* and also encoded in *A. laidlawii* [132–134].

Acetyl CoA can be converted to acetyl phosphate phosphotransacetylase (Pta) in many mycoplasmas. Subsequently, acetyl phosphate can be transformed by acetate kinase (AckA) to acetate and ATP (Figure 9). However, Pta, which is encoded on many mycoplasma chromosomes close to *ackA*, is absent from all four phytoplasma genomes [14, 20]. Lacking Pta could be substituted by another phosphotransacetylase (PduL) described for *Salmonella*

*enterica* subsp. *enterica* serovar typhimurium (Acc. no. AAD39011), since it is encoded in all four phytoplasma chromosomes. This assignment of PduL is in accordance with the kinetic parameter with a KM value of 0.97 mM for acetyl phosphate and a V<sub>max</sub> of 13.4  $\mu$ M/min/mg enzyme with acetyl phosphate as substrate (according to Uniprot entry Q9XDN5).

The assignment of the phosphotransacetylase as propanediol utilisation-like protein PduL in phytoplasmas remains ambiguous. Propanediol is produced by fermentation of the common plant sugars rhamnose and fucose. In *S. enterica*, PduL is part of the *pdu*-operon and involved in the coenzyme-B<sub>12</sub>-dependent degradation of 1,2-propanediol [135–137]. It seems to be likely that in phytoplasmas, phosphotransacetylase PduL does not function in propanediol degradation, since other genes of this pathway (e.g., *pduCDE*, *pduQ*, *pduP*, and *pduW*) are absent in phytoplasmas and *A. laidlawii* (except for *pduL*). Thus, PduL could use acetyl-CoA instead of propionyl-CoA as substrate.

### 8.3. Associated Processes

**8.3.1. Proteins Involved in NAD Synthesis.** All four phytoplasma chromosomes encode the glutamine-dependent NAD<sup>+</sup> synthetase NadE (ST49). In this ATP-dependent process nicotinate adenine dinucleotide (deamido-NAD<sup>+</sup>), L-glutamine, and H<sub>2</sub>O are used to form L-glutamate, AMP, diphosphate, H<sup>+</sup>, and NAD<sup>+</sup>. Other proteins involved in the generation of NAD<sup>+</sup> were not identified in the phytoplasma genomes. In contrast, *A. laidlawii* contains NadD and PncB allowing the formation of NAD<sup>+</sup> from nicotinate. It appears

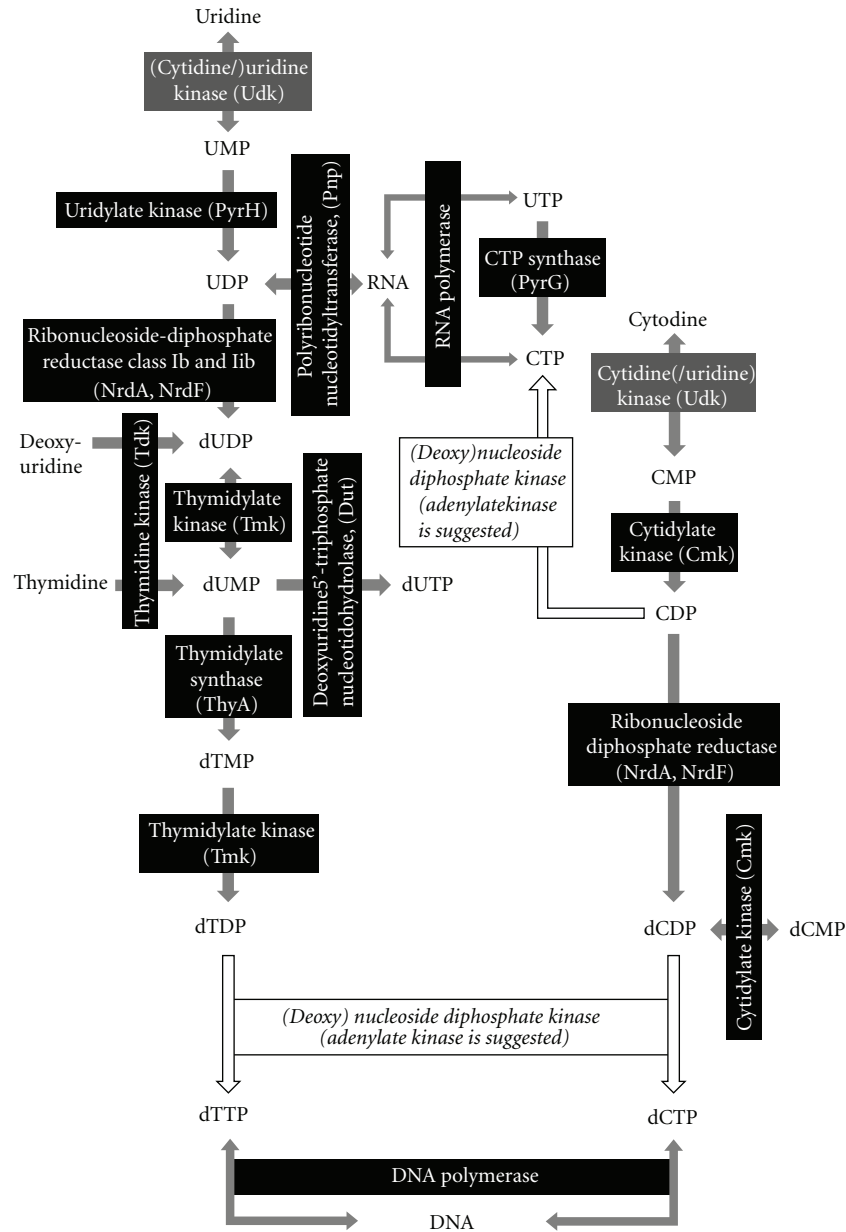


FIGURE 10: Principle of pyrimidine metabolism in phytoplasmas.

likely that phytoplasmas import deamido-NAD<sup>+</sup> or a precursor from their environment.

**8.3.2. Proteins Involved in Oxygen Detoxification.** All four genomes encode a superoxide dismutase (SOD), which converts O<sub>2</sub><sup>-</sup> to the less toxic H<sub>2</sub>O<sub>2</sub> (ST40). SOD activity is documented for several *Mollicutes* including the genera *Mycoplasma*, *Ureaplasma*, and *Acholeplasma* [138]. SOD requires metal cations such as copper, manganese, iron, or nickel. Prokaryotes were assigned to form cytoplasmatic MnSOD and/or FeSOD [139], but for *M. hyopneumoniae* the production of Cu/ZnSOD was shown [140]. The formation of MnSOD is reported for the acholeplasmas [141]. The catalase dismutase responsible for the conversion of H<sub>2</sub>O<sub>2</sub>

to nontoxic products was not identified in phytoplasmas so far. This is also reported for several other mycoplasmas and may contribute to virulence of these organisms by release of reactive H<sub>2</sub>O<sub>2</sub> [142]. It remains unclear, whether the release of H<sub>2</sub>O<sub>2</sub> will weaken or damage the plant host upon presence of phytoplasmas inside the sieve elements.

**8.3.3. Lipoyl-Protein Ligase.** Lipoic acid derivatives act as cofactors in enzymatic systems, such as the pyruvate dehydrogenase [143] in the phytoplasmas. All four chromosomes encode the ATP-dependent lipoyl-protein ligase LplA (ST50), which preferentially utilizes imported lipoate to form lipoyl adenylate or an octanoylated protein from octanoylate [144].



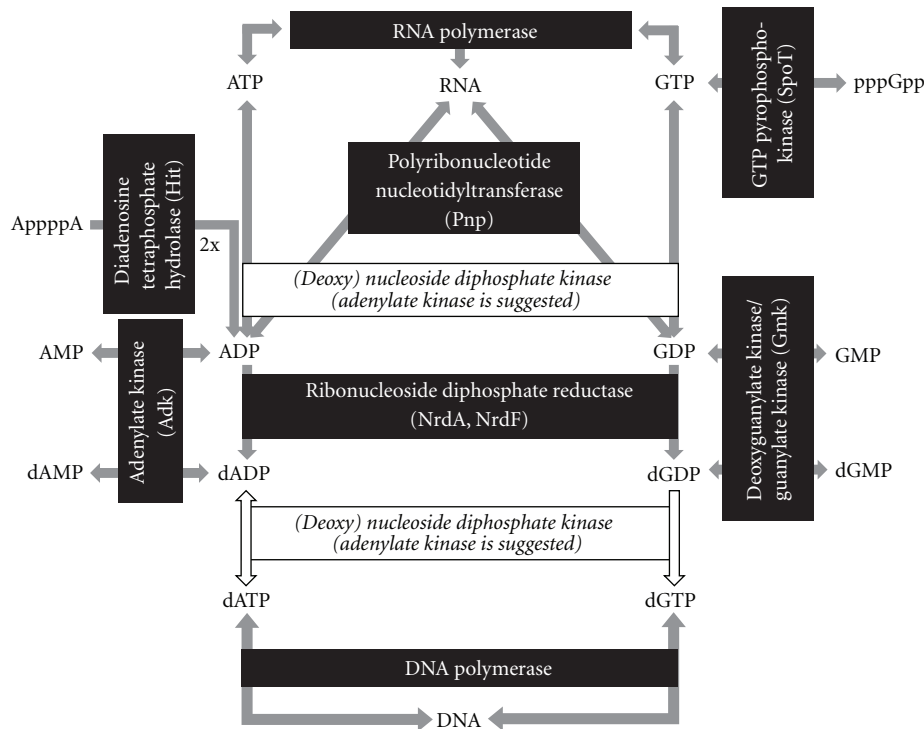


FIGURE 11: Principle of purine metabolism in phytoplasmas.

## 9. Purine and Pyrimidine Synthesis

Purine and pyrimidine metabolism in *Mollicutes* was reviewed recently [145]. Therefore, the following focuses on the genetic key elements or aspects of general interest for the phytoplasmas (ST42). Since phytoplasmas possess no genetic repertoire for *de novo* synthesis of purine or pyrimidine bases [48], they depend like many other mollicutes on environmentally derived nucleotide precursors [145]. However, nucleobase or nucleoside transporters were not identified in phytoplasmas so far [48], but it was suggested for mycoplasmas that the limited repertoire of transporters could be tuned to a wider variety of substrates [146]. Membrane-associated nucleases are suggested for the nucleotide precursor uptake and were identified in 20 mycoplasma species [147]. Membrane nucleases such as MnuA of *M. pulmonis* [147] were characterized, and putative orthologs were identified in *M. hyopneumoniae*, *M. gal-lisepticum*, *M. pneumoniae*, *M. penetrans*, and *U. urealyticum* [145]. However, it was not possible to identify candidates in phytoplasmas so far. The uptake and incorporation of dNMP without prior dephosphorylation was shown for *M. mycoides* subsp. *mycoides* [148]. The presence of such a strategy in phytoplasmas is important for nucleotide and energy metabolism. However, all four genomes encode common pathways for purine and pyrimidine synthesis with one exception (Figure 10). The usage of uridine and cytosine in the 'Ca. P. australiense' and 'Ca. P. mali' remains questionable, because the cytosine/uridine kinase (Udk) mediating the formation of CMP/UMP is absent in both chromosomes. However, it is present in *A. laidlawii* (YP\_001620379).

AMP and dAMP as well as GMP and dGMP represent the entry points into purine metabolism in all four genomes (Figure 11). The (deoxy)nucleoside diphosphate kinase (Ndk) is absent in the four phytoplasma genomes. This was also observed with many other mycoplasmas, the 6-phosphofructo-, phosphoglycerate-, pyruvate-, and acetate-kinases which could use other ribo- and deoxyribopurine and pyrimidine NDPs and NTPs besides ADP/ATP [146]. Similarly it was shown for *Mycobacterium tuberculosis* that the adenylate kinase (Adk) acts as general (deoxy)nucleoside diphosphate kinase (NdK) [149], a scenario that was also proposed for mycoplasmas [145].

CMP and CTP are also produced as byproducts within the phospholipid metabolism. However, it should be also noted that such function of the Adk might also result in an imbalance of the nucleotide pool and in the low G + C content of phytoplasma chromosomes in evolution. Notably, 'Ca. P. mali' with the lowest G + C content encodes one *adk* on each terminal repeat of the chromosome.

## 10. Miscellaneous

**10.1. Lipid Synthesis.** All four phytoplasma genomes encode a common biosynthetic pathway for essential phospholipids (Figure 12, ST44), that is, for the formation of CDP-diacylglycerol from an acyl phosphate and dihydroxyacetone phosphate. Moreover, the genetic modules for the generation of L-1-phosphatidyl-glycerol and L-1-phosphatidyl-ethanolamine are present. The potential to form cardiolipin from L-1-phosphatidyl-glycerol remains unclear, since a cardiolipin synthase is apparently not encoded.

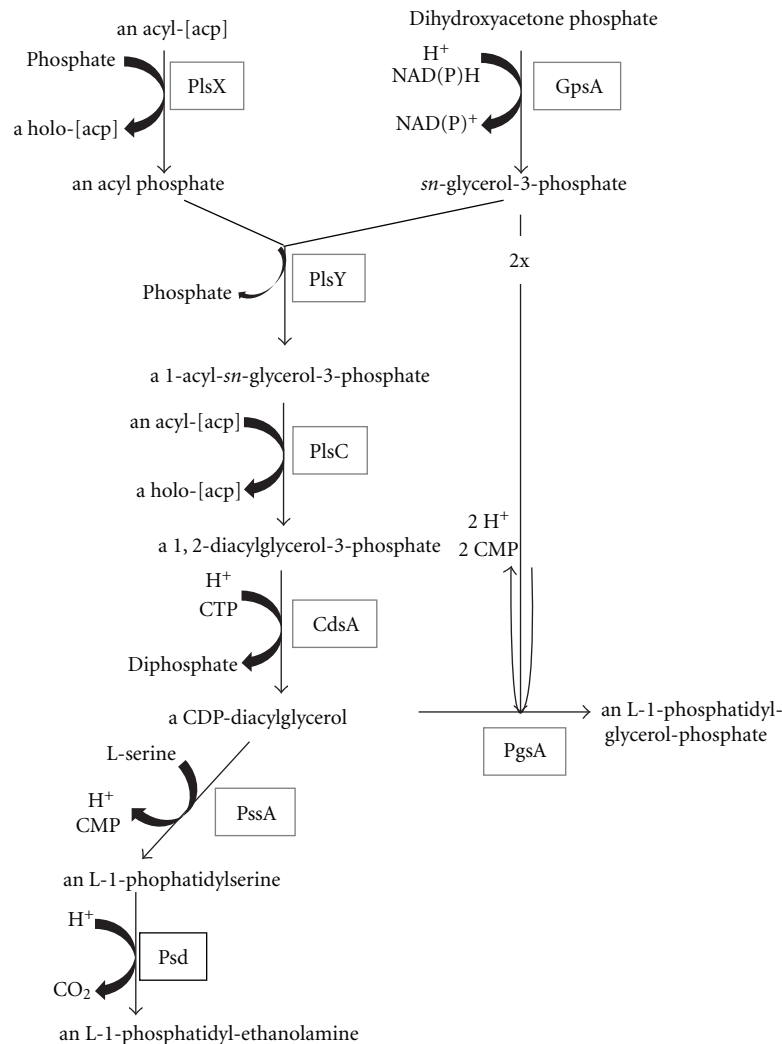


FIGURE 12: Key elements of the lipid metabolism.

**10.2. Proteolysis.** Diverse proteases involved in the breakdown of proteins are encoded in the phytoplasma genomes. Several encoded peptidases enable the degradation of essential imported peptides and regulation of intracellular products.

A remarkable number of zinc-dependent proteases are predicted from the phytoplasma genomes. For example the four genomes share the protease PmbA and Zn-dependent protease TldD (ST38), and the exact role in phytoplasmas remains elusive.

The annotated zinc-dependent HflB metalloproteases may play a prominent role in the '*Ca. P. asteris*', due to the high number of paralogs: 19 in strain OY-M and 10 in strain AY-WB. High numbers of paralogs are also present in '*Ca. P. australiense*' [6] and '*Ca. P. mali*' [11]. These *hflB* gene products show differences in the encoded peptide length and domain composition. In *E. coli*, HflB is affiliated with multiple cellular functions, including a putative involvement in lysogeny of lambda phage [150]. Thus, one may speculate that the high number of *hflB* genes resulted from a high pressure of phage attacks. Studies on the various *hflB* genes

and their presence within the genomes of differently virulent (mild to severe) strains are on the way.

**10.3. Amino Acid Synthesis and Modification.** It was not possible to detect proteins involved in the synthesis of amino acids, except for S-adenosyl-L-methionine (MetK), L-asparagine (AsnB), and L-ornithine (ArgE) (ST41, supplementary Figure 1 (SF1)). However, these enzymes are neither encoded in all four genomes nor do they enable the phytoplasmas to synthesise any amino acid [48]. Thus, all necessary amino acids or peptides have to be imported.

**10.4. Riboflavin Synthesis.** The only detected putative protein involved in riboflavin syntheses was ATP-dependent riboflavin kinase (ST45) in strain OY-M.

**10.5. Folate Synthesis.** Folic acid (or vitamin B9) is a precursor of the coenzyme tetrahydrofolate (THF). All four phytoplasma genomes (ST46) encode dihydrofolate reductase (FolA), which uses 7,8-dihydrofolate monoglutamate

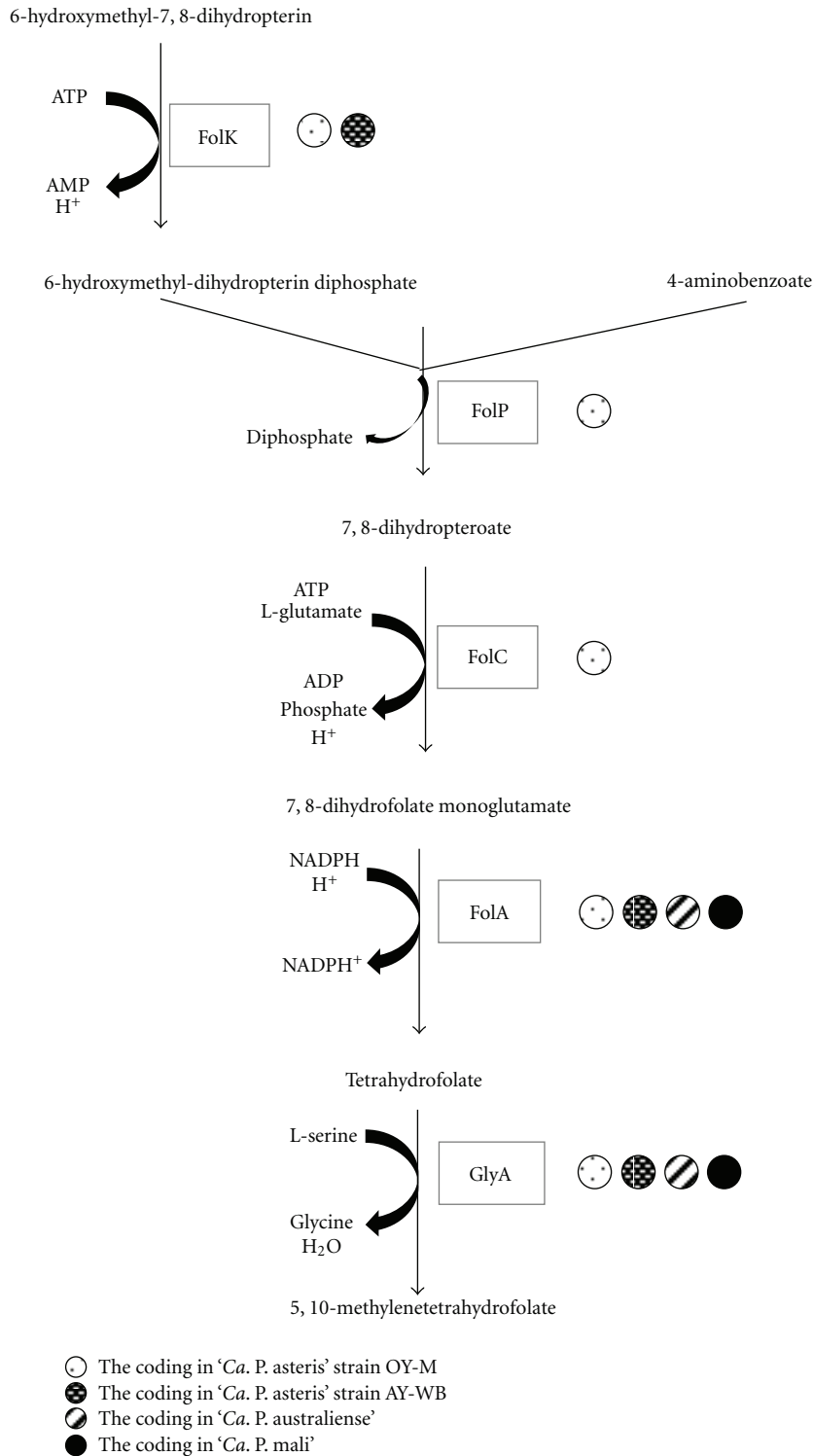


FIGURE 13: Folate synthesis.

and NADPH + H<sup>+</sup> for the generation of THF (Figure 13). Only strain OY-M encodes a gene set, which may allow generating THF from other precursors [49]. In the clover phyllody phytoplasma, genes assigned as *folP* and *folK* encode frameshifts and represent pseudogenes

[151]. It remains unclear how phytoplasmas obtain 7,8-dihydrofolate monoglutamate and how they process 5,10-methylenetetrahydrofolate. The incomplete folate synthesis may indicate the loss of this genetic modules and folate-dependance on the host. This may influence the host because

folate is also involved in photorespiration, amino acid metabolism, and chloroplastic protein biosynthesis in plants [152].

**10.6. Thiamine Synthesis.** All four phytoplasma genomes encode the ATP-dependent thiamine-phosphate phosphotransferase ThiJ (syn. ThiL) (ST47), which forms the coenzyme thiamine pyrophosphate (syn. thiamine diphosphate) from thiamine phosphate. ThiI, which is encoded in the chromosomes except for '*Ca. P. mali*', is required for thiazole synthesis in the thiamine biosynthesis pathway [153].

**10.7. Pyridoxal phosphate Synthesis.** The kinase PdxK of the vitamin B6 metabolism is only encoded in the genome of '*Ca. P. asteris*' OY-M (ST48). This ATP-dependent kinase catalyses the formation of pyridoxine-5'-phosphate from pyridoxal, pyridoxine, and pyridoxamine.

**10.8. Iron-Sulfur Cluster Biosynthesis.** NifU-like proteins are predicted to occur in all four phytoplasmas (ST39). Proteins such as NifS and NifU are required for the formation of metal clusters of nitrogenase in *Azotobacter vinelandii* and the maturation of other FeS proteins. No further genes associated with nitrogen fixation were found in the phytoplasma genomes. Notably, *nifU*-like genes are also encoded in genomes of other organisms lacking the ability to fix nitrogen [154].

**10.9. Acyl Carrier Protein Metabolism.** Two proteins assigned to the acyl carrier metabolism are encoded in all four genomes (ST43): the acyl carrier protein AcpP, which is involved in the biosynthesis of fatty acids and membrane-derived oligosaccharides in *E. coli* [155] and the holo-[ACP] synthase (AcpS) involved in carrier formation in lipid synthesis [156].

**10.10. Phosphate Metabolism.** A set of six proteins involved in the metabolism of phosphorous compounds is encoded in all four genomes (ST51). The inorganic pyrophosphatases (Ppa) catalyse the hydrolysis of inorganic pyrophosphate (PPi) to two orthophosphates (Pi), regulating the PPi pool which is replenished by various metabolic processes [157]. Additional predicted proteins related to phosphate metabolism are phosphohydrolases, metallophosphoesterases, and lysophospholipases.

**10.11. beta-Glucanase.** Several genes with a weak catalytic or unclear pathway assignment are shared between some of the four phytoplasmas. An example is the predicted endoglucanase FrvX (M42 peptidase family). Considering that endo-1,4-beta-glucanase proteins in *Arabidopsis thaliana* are associated with plant growth (in particular elongation), xylem development, and cell wall thickening [158] the phytoplasma FrvX protein could potentially contribute to virulence (ST52).

## 11. Outlook

Future research has to experimentally validate the functions of the key proteins in phytoplasmas. The labelling of

substrates and heterologous expression in cultivable hosts are possible strategies as long as axenic cultures of phytoplasmas are unavailable. Beside this basic molecular research at least two further fields of research have to be developed for phytoplasmas.

First, a '*Ca. Phytoplasma species*' with a less reduced genome and/or a '*Ca. Phytoplasma species*' closer affiliating with the phylogenetic separation of *Acholeplasma* and '*Ca. Phytoplasma*' has to be identified. Studies on such an ancestor will provide insights into the evolution of the genomes and the host relationship.

Second, gene expression data from a strain with complete and a host with at least a draft genome sequence are needed. To date, most insights were obtained from pathogens, but these data need to be integrated with the situation in the healthy plant, to learn more about the overall plant response.

## Acknowledgment

This work was supported by COST Action FA0807 (Integrated Management of Phytoplasma Epidemics in Different Crop Systems) and the Deutsche Forschungsgemeinschaft through Projects KU 2679/2-1 and BU 890/21-1.

## References

- [1] R. E. McCoy, A. Caudwell, C. J. Chang et al., "Plant diseases associated with mycoplasma-like organisms," in *The Mycoplasmas*, R. F. Whitcomb and J. G. Tully, Eds., vol. 5, pp. 546–640, Academic Press, New York, NY, USA, 1989.
- [2] E. Strauss, "Phytoplasma research begins to bloom," *Science*, vol. 325, no. 5939, pp. 388–390, 2009.
- [3] Y. T. M. Doi, K. Yora, and H. Asuyama, "Mycoplasma or PLT-group-like organisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows or paulownia witches broom," *Annals of the Phytopathological Society of Japan*, vol. 33, pp. 259–266, 1967.
- [4] G. Firrao, M. Andersen, A. Bertaccini et al., "'*Candidatus Phytoplasma*', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects," *International Journal of Systematic and Evolutionary Microbiology*, vol. 54, no. 4, pp. 1243–1255, 2004.
- [5] D. Harrison and R. E. Davis, "Genus I. '*Candidatus Phytoplasma*' gen. nov. IRPCM phytoplasma/spiroplasma working team 2004, 1244," in *The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*, N. R. Krieg, W. Ludwig, W. B. Whitman et al., Eds., vol. 4, pp. 696–718, Springer, New York, NY, USA, 2011.
- [6] R. E. Davis et al., "'*Candidatus Phytoplasma sudamericanum*', a novel taxon, and strain PassWB-Br4, a new subgroup 16SrIII-V phytoplasma, from diseased passion fruit (*Passiflora edulis* f. *flavicarpa* Deg.)," *International Journal of Systematic and Evolutionary Microbiology*. In press.
- [7] I. M. Lee, K. D. Bottner-Parker, Y. Zhao, W. Villalobos, and L. Moreira, "'*Candidatus Phytoplasma costaricanum*' a new phytoplasma associated with a newly emerging disease in soybean in Costa Rica," *International Journal of Systematic and Evolutionary Microbiology*, vol. 61, no. 12, pp. 2822–2826, 2011.

- [8] I. M. Lee, R. E. Davis, and D. E. Gundersen-Rindal, "Phytoplasma: phytopathogenic mollicutes," *Annual Review of Microbiology*, vol. 54, pp. 221–255, 2000.
- [9] H. Neimark, *Phylogenetic Relationships between Mycoplasmas and Other Prokaryotes*, Academic Press, New York, NY, USA, 1979.
- [10] C. R. Woese, J. Maniloff, and L. B. Zablen, "Phylogenetic analysis of the mycoplasmas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 1, pp. 494–498, 1980.
- [11] M. J. Rogers, J. Simmons, and R. T. Walker, "Construction of the mycoplasma evolutionary tree from 5S rRNA sequence data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 82, no. 4, pp. 1160–1164, 1985.
- [12] P. O. Lim and B. B. Sears, "Evolutionary relationships of a plant-pathogenic mycoplasma-like organism and *Acholeplasma laidlawii* deduced from two ribosomal protein gene sequences," *Journal of Bacteriology*, vol. 174, no. 8, pp. 2606–2611, 1992.
- [13] K. F. Toth, N. Harrison, and B. B. Sears, "Phylogenetic relationships among members of the class Mollicutes deduced from rps3 gene sequences," *International Journal of Systematic Bacteriology*, vol. 44, no. 1, pp. 119–124, 1994.
- [14] S. Razin, D. Yogeve, and Y. Naot, "Molecular biology and pathogenicity of mycoplasmas," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 4, pp. 1094–1156, 1998.
- [15] W. G. Weisburg, J. G. Tully, D. L. Rose et al., "A phylogenetic analysis of the mycoplasmas: basis for their classification," *Journal of Bacteriology*, vol. 171, no. 12, pp. 6455–6467, 1989.
- [16] P. Carle, F. Laigret, J. G. Tully, and J. M. Bove, "Heterogeneity of genome sizes within the genus *Spiroplasma*," *International Journal of Systematic Bacteriology*, vol. 45, no. 1, pp. 178–181, 1995.
- [17] H. Neimark and B. C. Kirkpatrick, "Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic mycoplasma-like organisms," *Molecular Microbiology*, vol. 7, no. 1, pp. 21–28, 1993.
- [18] C. Marcone and E. Seemüller, "A chromosome map of the European stone fruit yellows phytoplasma," *Microbiology*, vol. 147, no. 5, pp. 1213–1221, 2001.
- [19] C. Marcone, H. Neimark, A. Ragozzino, U. Lauer, and E. Seemüller, "Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups," *Phytopathology*, vol. 89, no. 9, pp. 805–810, 1999.
- [20] M. Kube, B. Schneider, H. Kuhl et al., "The linear chromosome of the plant-pathogenic mycoplasma "*Candidatus* *Phytoplasma mali*"", *BMC Genomics*, vol. 9, article 306, 2008.
- [21] N. M. Christensen, M. Nicolaisen, M. Hansen, and A. Schulz, "Distribution of phytoplasmas in infected plants as revealed by real-time PCR and bioimaging," *Molecular Plant-Microbe Interactions*, vol. 17, no. 11, pp. 1175–1184, 2004.
- [22] K. Oshima, T. Shiomi, T. Kuboyama et al., "Isolation and characterization of derivative lines of the onion yellows phytoplasma that do not cause stunting or phloem hyperplasia," *Phytopathology*, vol. 91, no. 11, pp. 1024–1029, 2001.
- [23] D. R. Webb, R. G. Bonfiglioli, L. Carraro, R. Osler, and R. H. Symons, "Oligonucleotides as hybridization probes to localize phytoplasmas in host plants and insect vectors," *Phytopathology*, vol. 89, no. 10, pp. 894–901, 1999.
- [24] K. J. Oparka and R. Turgeon, "Sieve elements and companion cells—traffic control centers of the phloem," *Plant Cell*, vol. 11, no. 4, pp. 739–750, 1999.
- [25] R. D. Sjölund, "The phloem sieve element: a river runs through it," *Plant Cell*, vol. 9, no. 7, pp. 1137–1146, 1997.
- [26] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.
- [27] S. Dinant and R. Lemoine, "The phloem pathway: new issues and old debates," *Comptes Rendus—Biologies*, vol. 333, no. 4, pp. 307–319, 2010.
- [28] M. van Helden, W. F. Tjallingii, and T. A. van Beek, "Phloem sap collection from lettuce (*Lactuca sativa* L.): chemical comparison among collection methods," *Journal of Chemical Ecology*, vol. 20, no. 12, pp. 3191–3206, 1994.
- [29] A. J. E. van Bel and P. H. Hess, "Hexoses as phloem transport sugars: the end of a dogma?" *Journal of Experimental Botany*, vol. 59, no. 2, pp. 261–272, 2008.
- [30] S. Dinant, J. L. Bonnemain, C. Girousse, and J. Kehr, "Phloem sap intricacy and interplay with aphid feeding," *Comptes Rendus—Biologies*, vol. 333, no. 6-7, pp. 504–515, 2010.
- [31] N. Noiraud, L. Maurousset, and R. Lemoine, "Identification of a mannitol transporter, AgMaT1, in celery phloem," *Plant Cell*, vol. 13, no. 3, pp. 695–705, 2001.
- [32] M. Ramsperger-Gleixner, D. Geiger, R. Hedrich, and N. Sauer, "Differential expression of sucrose transporter and polyol transporter genes during maturation of common plantain companion cells," *Plant Physiology*, vol. 134, no. 1, pp. 147–160, 2004.
- [33] M. Rokitta, A. D. Peuke, U. Zimmermann, and A. Haase, "Dynamic studies of phloem and xylem flow in fully differentiated plants by fast nuclear-magnetic-resonance microimaging," *Protoplasma*, vol. 209, no. 1-2, pp. 126–131, 1999.
- [34] L. A. Cernusak, D. J. Arthur, J. S. Pate, and G. D. Farquhar, "Water relations link carbon and oxygen isotope discrimination to phloem sap sugar concentration in *Eucalyptus globulus*," *Plant Physiology*, vol. 131, no. 4, pp. 1544–1554, 2003.
- [35] H. Ziegler, "Nature of substances in phloem," in *Encyclopedia of Plant Physiology, Transport in Plants*, M. H. Zimmermann and J. A. Milburn, Eds., vol. 1, pp. 57–100, Springer, Berlin, Germany, 1975.
- [36] A. Gojon, P. Nacry, and J. C. Davidian, "Root uptake regulation: a central process for NPS homeostasis in plants," *Current Opinion in Plant Biology*, vol. 12, no. 3, pp. 328–338, 2009.
- [37] T. Y. Liu, C. Y. Chang, and T. J. Chiou, "The long-distance signaling of mineral macronutrients," *Current Opinion in Plant Biology*, vol. 12, no. 3, pp. 312–319, 2009.
- [38] J. Kehr and A. Buhtz, "Long distance transport and movement of RNA through the phloem," *Journal of Experimental Botany*, vol. 59, no. 1, pp. 85–92, 2008.
- [39] B. D. Pant, M. Musialak-Lange, P. Nuc et al., "Identification of nutrient-responsive Arabidopsis and rapeseed microRNAs by comprehensive real-time polymerase chain reaction profiling and small RNA sequencing," *Plant Physiology*, vol. 150, no. 3, pp. 1541–1555, 2009.
- [40] M. K. Lin, Y. J. Lee, T. J. Lough, B. S. Phimmey, and W. J. Lucas, "Analysis of the pumpkin phloem proteome provides insights into angiosperm sieve tube function," *Molecular and Cellular Proteomics*, vol. 8, no. 2, pp. 343–356, 2009.
- [41] P. Jones, "Phytoplasma plant pathogens," in *Plant Pathologist's Pocketbook*, J. M. Waller, J. M. Lenné, and S. J. Waller, Eds., CABI, New York, NY, USA, 3rd edition, 2002.
- [42] P. G. Weintraub and L. Beanland, "Insect vectors of phytoplasmas," *Annual Review of Entomology*, vol. 51, pp. 91–111, 2006.

- [43] C. J. Mayer, A. Vilcinskas, and J. Gross, "Pathogen-induced release of plant allomone manipulates vector insect behavior," *Journal of Chemical Ecology*, vol. 34, no. 12, pp. 1518–1522, 2008.
- [44] B. C. Kirkpatrick, D. C. Stenger, T. Jack Morris, and A. H. Purcell, "Cloning and detection of DNA from a nonculturable plant pathogenic mycoplasma-like organism," *Science*, vol. 238, no. 4824, pp. 197–200, 1987.
- [45] N. Moriwaki, K. Matsushita, M. Nishina, and Y. Kono, "High concentrations of trehalose in aphid hemolymph," *Applied Entomology and Zoology*, vol. 38, no. 2, pp. 241–248, 2003.
- [46] L. Galetto, D. Bosco, R. Balestrini, A. Genre, J. Fletcher, and C. Marzachi, "The major antigenic membrane protein of *Candidatus Phytoplasma asteris*" selectively interacts with ATP synthase and actin of leafhopper vectors," *PLoS One*, vol. 6, no. 7, Article ID e22571, 2011.
- [47] J. P. Gourret, P. L. Maillat, and J. Gouranton, "Virus-like particles associated with the mycoplasmas of clover phyllody in the plant and in the insect vector," *Journal of General Microbiology*, vol. 74, no. 2, pp. 241–249, 1973.
- [48] K. Oshima, S. Kakizawa, H. Nishigawa et al., "Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma," *Nature Genetics*, vol. 36, no. 1, pp. 27–29, 2004.
- [49] X. Bai, J. Zhang, A. Ewing et al., "Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts," *Journal of Bacteriology*, vol. 188, no. 10, pp. 3682–3696, 2006.
- [50] L. T. T. Tran-Nguyen, M. Kube, B. Schneider, R. Reinhardt, and K. S. Gibb, "Comparative genome analysis of *Candidatus Phytoplasma australiense*" (subgroup tuf-Australia I; rp-A) and "Ca. phytoplasma asteris" strains OY-M and AY-WB," *Journal of Bacteriology*, vol. 190, no. 11, pp. 3979–3991, 2008.
- [51] R. D. Fleischmann, M. D. Adams, O. White et al., "Whose genome random sequencing and assembly of *Haemophilus influenzae* Rd," *Science*, vol. 269, no. 5223, pp. 496–521, 1995.
- [52] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyström, "Real-time DNA sequencing using detection of pyrophosphate release," *Analytical Biochemistry*, vol. 242, no. 1, pp. 84–89, 1996.
- [53] L. T. T. Tran-Nguyen and K. S. Gibb, "Extrachromosomal DNA isolated from tomato big bud and *Candidatus Phytoplasma australiense* phytoplasma strains," *Plasmid*, vol. 56, no. 3, pp. 153–166, 2006.
- [54] H. Nishigawa, K. Oshima, S. Kakizawa et al., "A plasmid from a non-insect-transmissible line of a phytoplasma lacks two open reading frames that exist in the plasmid from the wild-type line," *Gene*, vol. 298, no. 2, pp. 195–201, 2002.
- [55] G. C. Chaconas and C. W. Chen, "Replication of linear bacterial chromosomes: no longer going around the circle," in *The Bacterial Chromosome*, N. Higgins, Ed., pp. 525–539, ASM Press, Washington, DC, USA, 2005.
- [56] W. Wei, R. E. Davis, R. Jomantieni, and Y. Zhao, "Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 33, pp. 11827–11832, 2008.
- [57] M. Bose and R. D. Barber, "Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences," *In Silico Biology*, vol. 6, no. 3, pp. 223–227, 2006.
- [58] A. C. Padovan, G. Firrao, B. Schneider, and K. S. Gibb, "Chromosome mapping of the sweet potato little leaf phytoplasma reveals genome heterogeneity within the phytoplasmas," *Microbiology*, vol. 146, no. 4, part 4, pp. 893–902, 2000.
- [59] T. Y. Toruño, M. Seruga Musić, S. Simi, M. Nicolaisen, and S. A. Hogenhout, "Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts," *Molecular Microbiology*, vol. 77, no. 6, pp. 1406–1415, 2010.
- [60] S. A. Hogenhout and S. A. Musić, "Phytoplasma genomics, from sequencing to comparative and functional genomics—what have we learnt?" in *Phytoplasmas: Genomes, Plant Hosts and Vectors*, P. G. Weintraub and P. Jones, Eds., pp. 19–36, CABI, Wallingford, UK, 2010.
- [61] Y. Xiong and T. H. Eickbush, "Origin and evolution of retroelements based upon their reverse transcriptase sequences," *The EMBO Journal*, vol. 9, no. 10, pp. 3353–3362, 1990.
- [62] R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V. Koonin, "The COG database: a tool for genome-scale analysis of protein functions and evolution," *Nucleic Acids Research*, vol. 28, no. 1, pp. 33–36, 2000.
- [63] J. M. Kaguni, L. L. Bertsch, D. Bramhill et al., "Initiation of replication of the *Escherichia coli* chromosomal origin reconstituted with purified enzymes," *Basic Life Sciences*, vol. 30, pp. 141–150, 1985.
- [64] T. A. Baker, B. E. Funnell, and A. Kornberg, "Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin in vitro," *The Journal of Biological Chemistry*, vol. 262, no. 14, pp. 6877–6885, 1987.
- [65] W. Seufert, B. Dobrinski, R. Lurz, and W. Messer, "Functionality of the dnaA protein binding site in DNA replication is orientation-dependent," *The Journal of Biological Chemistry*, vol. 263, no. 6, pp. 2719–2723, 1988.
- [66] C. S. McHenry, "Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences," *Molecular Microbiology*, vol. 49, no. 5, pp. 1157–1165, 2003.
- [67] J. R. Lobry, "Origin of replication of *Mycoplasma genitalium*," *Science*, vol. 272, no. 5262, pp. 745–746, 1996.
- [68] J. R. Lobry, "Asymmetric substitution patterns in the two DNA strands of bacteria," *Molecular Biology and Evolution*, vol. 13, no. 5, pp. 660–665, 1996.
- [69] A. C. Frank and J. R. Lobry, "Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes," *Bioinformatics*, vol. 16, no. 6, pp. 560–561, 2000.
- [70] C. M. M. Cordova, C. Lartigue, P. Sirand-Pugnet, J. Renaudin, R. A. F. Cunha, and A. Blanchard, "Identification of the origin of replication of the *Mycoplasma pulmonis* chromosome and its use in oriC replicative plasmids," *Journal of Bacteriology*, vol. 184, no. 19, pp. 5426–5435, 2002.
- [71] R. S. Fuller, B. E. Funnell, and A. Kornberg, "The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites," *Cell*, vol. 38, no. 3, pp. 889–900, 1984.
- [72] D. S. S. Hudspeth and P. S. Vary, "spoVG sequence of *Bacillus megaterium* and *Bacillus subtilis*," *Biochimica et Biophysica Acta*, vol. 1130, no. 2, pp. 229–231, 1992.
- [73] J. I. Glass, N. Assad-Garcia, N. Alperovich et al., "Essential genes of a minimal bacterium," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 2, pp. 425–430, 2006.
- [74] L. N. Kinch, K. Ginalski, L. Rychlewski, and N. V. Grishin, "Identification of novel restriction endonuclease-like fold

- families among hypothetical proteins," *Nucleic Acids Research*, vol. 33, no. 11, pp. 3598–3605, 2005.
- [75] H. I. Miller and D. I. Friedman, "An E. coli gene product required for  $\lambda$  site-specific recombination," *Cell*, vol. 20, no. 3, pp. 711–719, 1980.
- [76] D. N. Arnosti and M. J. Chamberlin, "Secondary  $\sigma$  factor controls transcription of flagellar and chemotaxis genes in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 3, pp. 830–834, 1989.
- [77] A. F. González Barrios, R. Zuo, D. Ren, and T. K. Wood, "Hha, YbaJ, and OmpA regulate *Escherichia coli* K12 biofilm formation and conjugation plasmids abolish motility," *Biotechnology and Bioengineering*, vol. 93, no. 1, pp. 188–200, 2006.
- [78] D. F. Ward and M. E. Gottesman, "The nus mutations affect transcription termination in *Escherichia coli*," *Nature*, vol. 292, no. 5820, pp. 212–215, 1981.
- [79] J. Roberts and J. S. Park, "Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination," *Current Opinion in Microbiology*, vol. 7, no. 2, pp. 120–125, 2004.
- [80] J. W. Roberts, "Termination factor for RNA synthesis," *Nature*, vol. 224, no. 5225, pp. 1168–1174, 1969.
- [81] B. Schneider and E. Seemuller, "Presence of two sets of ribosomal genes in phytopathogenic mollicutes," *Applied and Environmental Microbiology*, vol. 60, no. 9, pp. 3409–3412, 1994.
- [82] K. Lagesen, P. Hallin, E. A. Rødland, H. H. Stærfeldt, T. Rognes, and D. W. Ussery, "RNAMmer: consistent and rapid annotation of ribosomal RNA genes," *Nucleic Acids Research*, vol. 35, no. 9, pp. 3100–3108, 2007.
- [83] T. M. Lowe and S. R. Eddy, "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence," *Nucleic Acids Research*, vol. 25, no. 5, pp. 955–964, 1997.
- [84] G. O. Bylund, L. C. Wipemo, L. A. C. Lundberg, and P. M. Wikström, "RimM and RbfA are essential for efficient processing of 16S rRNA in *Escherichia coli*," *Journal of Bacteriology*, vol. 180, no. 1, pp. 73–82, 1998.
- [85] Y. Qin, N. Polacek, O. Vesper et al., "The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome," *Cell*, vol. 127, no. 4, pp. 721–733, 2006.
- [86] G. Grentzmann, D. Brechemier-Baey, V. Heurgue-Hamard, and R. H. Buckingham, "Function of polypeptide chain release factor RF-3 in *Escherichia coli*. RF-3 action in termination is predominantly at UGA-containing stop signals," *The Journal of Biological Chemistry*, vol. 270, no. 18, pp. 10595–10600, 1995.
- [87] B. Bukau and A. L. Horwich, "The Hsp70 and Hsp60 chaperone machines," *Cell*, vol. 92, no. 3, pp. 351–366, 1998.
- [88] E. Laskowska, A. Wawrzynów, and A. Taylor, "IbpA and IbpB, the new heat-shock proteins, bind to endogenous *Escherichia coli* proteins aggregated intracellularly by heat shock," *Biochimie*, vol. 78, no. 2, pp. 117–122, 1996.
- [89] V. Tsilibaris, G. Maenhaut-Michel, and L. van Melderen, "Biological roles of the Lon ATP-dependent protease," *Research in Microbiology*, vol. 157, no. 8, pp. 701–713, 2006.
- [90] M. Carrión, M. J. Gómez, R. Merchante-Schubert, S. Don-garrá, and J. A. Ayala, "mraW, an essential gene at the dcw cluster of *Escherichia coli* codes for a cytoplasmic protein with methyltransferase activity," *Biochimie*, vol. 81, no. 8-9, pp. 879–888, 1999.
- [91] C. T. Parker, A. W. Kloser, C. A. Schnaitman, M. A. Stein, S. Gottesman, and B. W. Gibson, "Role of the rfaG and rfaP genes in determining the lipopolysaccharide core structure and cell surface properties of *Escherichia coli* K-12," *Journal of Bacteriology*, vol. 174, no. 8, pp. 2525–2538, 1992.
- [92] N. Umeda, T. Suzuki, M. Yukawa et al., "Mitochondria-specific RNA-modifying enzymes responsible for the biosynthesis of the wobble base in mitochondrial tRNAs: implications for the molecular pathogenesis of human mitochondrial diseases," *The Journal of Biological Chemistry*, vol. 280, no. 2, pp. 1613–1624, 2005.
- [93] D. Brégeon, V. Colot, M. Radman, and F. Taddei, "Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift," *Genes and Development*, vol. 15, no. 17, pp. 2295–2306, 2001.
- [94] K. von Meyenburg, B. B. Jorgensen, J. Nielsen, and F. G. Hansen, "Promoters of the atp operon coding for the membrane-bound ATP synthase of *Escherichia coli* mapped by Tn10 insertion mutations," *Molecular and General Genetics*, vol. 188, no. 2, pp. 240–248, 1982.
- [95] C. A. Hutchison, S. N. Peterson, S. R. Gill et al., "Global transposon mutagenesis and a minimal mycoplasma genome," *Science*, vol. 286, no. 5447, pp. 2165–2169, 1999.
- [96] M. J. Romanowski, J. B. Bonanno, and S. K. Burley, "Crystal structure of the *Escherichia coli* glucose-inhibited division protein B (GidB) reveals a methyltransferase fold," *Proteins*, vol. 47, no. 4, pp. 563–567, 2002.
- [97] A. Cimerman, D. Pacifico, P. Salar, C. Marzachi, and X. Foissac, "Striking diversity of vmp1, a variable gene encoding a putative membrane protein of the stolbur phytoplasma," *Applied and Environmental Microbiology*, vol. 75, no. 9, pp. 2951–2957, 2009.
- [98] S. Kakizawa, K. Oshima, and S. Namba, "Diversity and functional importance of phytoplasma membrane proteins," *Trends in Microbiology*, vol. 14, no. 6, pp. 254–256, 2006.
- [99] S. Suzuki, K. Oshima, S. Kakizawa et al., "Interaction between the membrane protein of a pathogen and insect microfilament complex determines insect-vector specificity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 11, pp. 4252–4257, 2006.
- [100] S. Kakizawa, K. Oshima, H. Y. Jung et al., "Positive selection acting on a surface membrane protein of the plant-pathogenic phytoplasmas," *Journal of Bacteriology*, vol. 188, no. 9, pp. 3424–3428, 2006.
- [101] S. A. Hogenhout, K. Oshima, E. D. Ammar, S. Kakizawa, H. N. Kingdom, and S. Namba, "Phytoplasmas: bacteria that manipulate plants and insects," *Molecular Plant Pathology*, vol. 9, no. 4, pp. 403–423, 2008.
- [102] S. Kakizawa, K. Oshima, Y. Ishii et al., "Cloning of immunodominant membrane protein genes of phytoplasmas and their in planta expression," *FEMS Microbiology Letters*, vol. 293, no. 1, pp. 92–101, 2009.
- [103] L. Käll, A. Krogh, and E. L. L. Sonnhammer, "A combined transmembrane topology and signal peptide prediction method," *Journal of Molecular Biology*, vol. 338, no. 5, pp. 1027–1036, 2004.
- [104] S. O. Kakizawa, K. Oshima, and S. Namba, "Functional genomics of phytoplasmas," in *Phytoplasmas: Genomes, Plant Hosts and Vectors*, P. J. Weintraub and P. Jones, Eds., pp. 37–50, CABI, Wallingford, UK, 2010.
- [105] X. Bai, V. R. Correa, T. Y. Toruño, E. D. Ammar, S. Kamoun, and S. A. Hogenhout, "AY-WB phytoplasma secretes a protein that targets plant cell nuclei," *Molecular Plant-Microbe Interactions*, vol. 22, no. 1, pp. 18–30, 2009.
- [106] A. Hoshi, K. Oshima, S. Kakizawa et al., "A unique virulence factor for proliferation and dwarfism in plants identified

- from a phytopathogenic bacterium," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 15, pp. 6416–6421, 2009.
- [107] S. Kakizawa, K. Oshima, H. Nishigawa et al., "Secretion of immunodominant membrane protein from onion yellows phytoplasma through the Sec protein-translocation system in *Escherichia coli*," *Microbiology*, vol. 150, no. 1, pp. 135–142, 2004.
- [108] J. C. Samuelson, M. Chen, F. Jiang et al., "YidC mediates membrane protein insertion in bacteria," *Nature*, vol. 406, no. 6796, pp. 637–641, 2000.
- [109] K. J. Pogliano and J. Beckwith, "Genetic and molecular characterization of the *Escherichia coli* secD operon and its products," *Journal of Bacteriology*, vol. 176, no. 3, pp. 804–814, 1994.
- [110] J. Fang and Y. Wei, "Expression, purification and Characterization of the *Escherichia coli* integral membrane protein YajC," *Protein and Peptide Letters*, vol. 18, no. 6, pp. 601–608, 2011.
- [111] H. Nielsen, J. Engelbrecht, S. Brunak, and G. von Heijne, "A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites," *International Journal of Neural Systems*, vol. 8, no. 5–6, pp. 581–599, 1997.
- [112] K. Nakai and M. Kanehisa, "Expert system for predicting protein localization sites in gram-negative bacteria," *Proteins*, vol. 11, no. 2, pp. 95–110, 1991.
- [113] R. Rabus, M. Kube, J. Heider et al., "The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1," *Archives of Microbiology*, vol. 183, no. 1, pp. 27–36, 2005.
- [114] H. G. Koch, T. Hengelage, C. Neumann-Haefelin et al., "In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein-targeting pathways of *Escherichia coli*," *Molecular Biology of the Cell*, vol. 10, no. 7, pp. 2163–2173, 1999.
- [115] H. G. Koch, M. Moser, and M. Müller, "Signal recognition particle-dependent protein targeting, universal to all kingdoms of life," *Reviews of Physiology, Biochemistry and Pharmacology*, vol. 146, pp. 55–94, 2003.
- [116] S. Brown, "4.5S RNA: does form predict function?" *New Biologist*, vol. 3, no. 5, pp. 430–438, 1991.
- [117] Z. Silva, M. M. Sampaio, A. Henne et al., "The high-affinity maltose/trehalose ABC transporter in the extremely thermophilic bacterium *Thermus thermophilus* HB27 also recognizes sucrose and palatinose," *Journal of Bacteriology*, vol. 187, no. 4, pp. 1210–1218, 2005.
- [118] A. André, W. Maccheroni, F. Doignon, M. Garnier, and J. Renaudin, "Glucose and trehalose PTS permeases of *Spiroplasma citri* probably share a single IIA domain, enabling the spiroplasma to adapt quickly to carbohydrate changes in its environment," *Microbiology*, vol. 149, no. 9, pp. 2687–2696, 2003.
- [119] D. Hekstra and J. Tommassen, "Functional exchangeability of the ABC proteins of the periplasmic binding protein-dependent transport systems Ugp and Mal of *Escherichia coli*," *Journal of Bacteriology*, vol. 175, no. 20, pp. 6546–6552, 1993.
- [120] S. Hunter, R. Apweiler, T. K. Attwood et al., "InterPro: the integrative protein signature database," *Nucleic Acids Research*, vol. 37, no. 1, pp. D211–D215, 2009.
- [121] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand, "Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms," *The Journal of Biological Chemistry*, vol. 277, no. 50, pp. 48949–48959, 2002.
- [122] I. Tobuereen Bots and H. Hagedorn, "Studies on the thiamine transport system in *Bacillus cereus*," *Archives of Microbiology*, vol. 113, no. 1–2, pp. 23–31, 1977 (German).
- [123] X. Foissac, J. L. Danet, C. Saillard et al., "Mutagenesis by insertion of Tn4001 into the genome of *Spiroplasma citri*: characterization of mutants affected in plant pathogenicity and transmission to the plant by the leafhopper vector *Circulifer haematocaps*," *Molecular Plant-Microbe Interactions*, vol. 10, no. 4, pp. 454–461, 1997.
- [124] S. I. Sukharev, P. Blount, B. Martinac, F. R. Blattner, and C. Kung, "A large-conductance mechanosensitive channel in *E. coli* encoded by mscL alone," *Nature*, vol. 368, no. 6468, pp. 265–268, 1994.
- [125] J. G. Voet and R. H. Abeles, "The mechanism of action of sucrose phosphorylase. Isolation and properties of a beta-linked covalent glucose-enzyme complex," *The Journal of Biological Chemistry*, vol. 245, no. 5, pp. 1020–1031, 1970.
- [126] R. Bruckner, E. Wagner, and F. Gotz, "Characterization of a sucrase gene from *Staphylococcus xylosus*," *Journal of Bacteriology*, vol. 175, no. 3, pp. 851–857, 1993.
- [127] P. J. Eastmond, Y. Li, and I. A. Graham, "Is trehalose-6-phosphate a regulator of sugar metabolism in plants?" *Journal of Experimental Botany*, vol. 54, no. 382, pp. 533–537, 2003.
- [128] Y. Zhang, L. F. Primavesi, D. Jhurrea et al., "Inhibition of SNF1-related protein kinase activity and regulation of metabolic pathways by trehalose-6-phosphate," *Plant Physiology*, vol. 149, no. 4, pp. 1860–1871, 2009.
- [129] M. J. Paul, D. Jhurrea, Y. Zhang et al., "Upregulation of biosynthetic processes associated with growth by trehalose 6-phosphate," *Plant Signaling and Behavior*, vol. 5, no. 4, pp. 386–392, 2010.
- [130] R. S. Stephens, S. Kalman, C. Lammel et al., "Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*," *Science*, vol. 282, no. 5389, pp. 754–759, 1998.
- [131] J. D. Pollack, M. V. Williams, and R. N. McElhaney, "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," *Critical Reviews in Microbiology*, vol. 23, no. 4, pp. 269–354, 1997.
- [132] G. Cohen and N. L. Somerson, "Glucose-dependent secretion and destruction of hydrogen peroxide by *Mycoplasma pneumoniae*," *Journal of Bacteriology*, vol. 98, no. 2, pp. 547–551, 1969.
- [133] G. Cohen and N. L. Somerson, "*Mycoplasma pneumoniae*: hydrogen peroxide secretion and its possible role in virulence," *Annals of the New York Academy of Sciences*, vol. 143, no. 1, pp. 85–87, 1967.
- [134] N. L. Somerson, B. E. Walls, and R. M. Chanock, "Hemolysin of *Mycoplasma pneumoniae*: tentative identification as a peroxidase," *Science*, vol. 150, no. 3693, pp. 226–227, 1965.
- [135] T. A. Bobik, G. D. Havemann, R. J. Busch, D. S. Williams, and H. C. Aldrich, "The propanediol utilization (pdu) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B12-dependent 1,2-propanediol degradation," *Journal of Bacteriology*, vol. 181, no. 19, pp. 5967–5975, 1999.



- [136] D. Walter, M. Ailion, and J. Roth, "Genetic characterization of the pdu operon: use of 1,2-propanediol in *Salmonella typhimurium*," *Journal of Bacteriology*, vol. 179, no. 4, pp. 1013–1022, 1997.
- [137] Y. Liu, N. A. Leal, E. M. Sampson, C. L. V. Johnson, G. D. Havemann, and T. A. Bobik, "PduL is an evolutionarily distinct phosphotransacylase involved in B 12-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar typhimurium LT2," *Journal of Bacteriology*, vol. 189, no. 5, pp. 1589–1596, 2007.
- [138] B. Meier and G. G. Habermehl, "Evidence for superoxide dismutase and catalase in mollicutes and release of reactive oxygen species," *Free Radical Research Communications*, vol. 12–13, pp. 451–454, 1991.
- [139] I. Fridovich, "Superoxide radical and superoxide dismutases," *Annual Review of Biochemistry*, vol. 64, pp. 97–112, 1995.
- [140] J. R. Chen, C. N. Weng, T. Y. Ho, I. C. Cheng, and S. S. Lai, "Identification of the copper-zinc superoxide dismutase activity in *Mycoplasma hyopneumoniae*," *Veterinary Microbiology*, vol. 73, no. 4, pp. 301–310, 2000.
- [141] S. J. O'Brien, J. M. Simonson, and M. W. Grabowski, "Analysis of multiple isoenzyme expression among twenty-two species of *Mycoplasma* and *Acholeplasma*," *Journal of Bacteriology*, vol. 146, no. 1, pp. 222–232, 1981.
- [142] B. Meier and G. G. Habermehl, "Evidence for superoxide dismutase and catalase in mollicutes and release of reactive oxygen species," *Archives of Biochemistry and Biophysics*, vol. 277, no. 1, pp. 74–79, 1990.
- [143] K. E. Reed and J. E. Cronan Jr., "Lipoic acid metabolism in *Escherichia coli*: sequencing and functional characterization of the lipA and lipB genes," *Journal of Bacteriology*, vol. 175, no. 5, pp. 1325–1336, 1993.
- [144] T. W. Morris, K. E. Reed, and J. E. Cronan Jr., "Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the lplA gene and gene product," *The Journal of Biological Chemistry*, vol. 269, no. 23, pp. 16091–16100, 1994.
- [145] C. V. Bizarro and D. C. Schuck, "Purine and pyrimidine nucleotide metabolism in Mollicutes," *Genetics and Molecular Biology*, vol. 30, no. 1, pp. 190–201, 2007.
- [146] J. D. Pollack, M. A. Myers, T. Dandekar, and R. Herrmann, "Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases," *OMICS A Journal of Integrative Biology*, vol. 6, no. 3, pp. 247–258, 2002.
- [147] F. C. Minion, K. J. Jarvill-Taylor, D. E. Billings, and E. Tigges, "Membrane-associated nuclease activities in *Mycoplasmas*," *Journal of Bacteriology*, vol. 175, no. 24, pp. 7842–7847, 1993.
- [148] G. A. M. Neale, A. Mitchell, and L. R. Finch, "Pathways of pyrimidine deoxyribonucleotide biosynthesis in *Mycoplasma mycoides* subsp. *mycoides*," *Journal of Bacteriology*, vol. 154, no. 1, pp. 17–22, 1983.
- [149] L. S. Meena, P. Chopra, R. S. Bedwal, and Y. Singh, "Nucleoside diphosphate kinase-like activity in adenylate kinase of *Mycobacterium tuberculosis*," *Biotechnology and Applied Biochemistry*, vol. 38, no. 2, pp. 169–174, 2003.
- [150] Y. Shotland, A. Shifrin, T. Ziv et al., "Proteolysis of bacteriophage lambda CII by *Escherichia coli* FtsH (HflB)," *Journal of Bacteriology*, vol. 182, no. 11, pp. 3111–3116, 2000.
- [151] R. E. Davis, R. Jomantiene, Y. Zhao, and E. L. Dally, "Folate biosynthesis pseudogenes, PsifolP and PsifolK, and an O-sialoglycoprotein endopeptidase gene homolog in the phytoplasma genome," *DNA and Cell Biology*, vol. 22, no. 11, pp. 697–706, 2003.
- [152] A. D. Hanson and J. F. Gregory, "Synthesis and turnover of folates in plants," *Current Opinion in Plant Biology*, vol. 5, no. 3, pp. 244–249, 2002.
- [153] E. Webb, K. Claas, and D. M. Downs, "Characterization of thiI, a new gene involved in thiazole biosynthesis in *Salmonella typhimurium*," *Journal of Bacteriology*, vol. 179, no. 13, pp. 4399–4402, 1997.
- [154] D. M. Hwang, A. Dempsey, K. T. Tan, and C. C. Liew, "A modular domain of NiFu, a nitrogen fixation cluster protein, is highly conserved in evolution," *Journal of Molecular Evolution*, vol. 43, no. 5, pp. 536–540, 1996.
- [155] M. Rawlings and J. E. Cronan Jr., "The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes," *The Journal of Biological Chemistry*, vol. 267, no. 9, pp. 5751–5754, 1992.
- [156] R. H. Lambalot and C. T. Walsh, "Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase," *The Journal of Biological Chemistry*, vol. 270, no. 42, pp. 24658–24661, 1995.
- [157] B. S. Cooperman, A. A. Baykov, and R. Lahti, "Evolutionary conservation of the active site of soluble inorganic pyrophosphatase," *Trends in Biochemical Sciences*, vol. 17, no. 7, pp. 262–266, 1992.
- [158] Z. Shani, M. Dekel, L. Roiz et al., "Expression of endo-1,4- $\beta$ -glucanase (cel1) in *Arabidopsis thaliana* is associated with plant growth, xylem development and cell wall thickening," *Plant Cell Reports*, vol. 25, no. 10, pp. 1067–1074, 2006.