LBNL-60246
Preprint

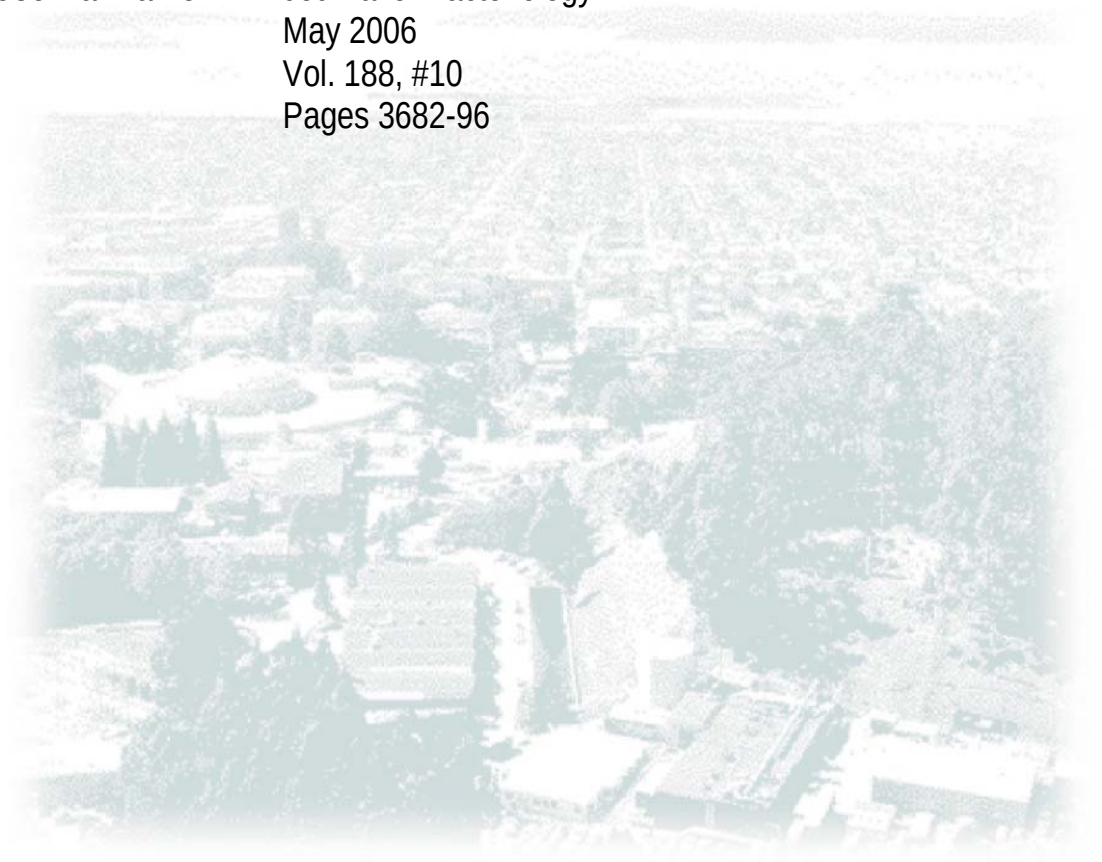
ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Title: Living with Genome Instability: the adaptation of Phytoplasmas to Diverse Environments of Their Insect and Plant Hosts

Author(s): Xiaodong Bai, Jianhua Zhang, et al

Division: Genomics

Journal Name: Journal of Bacteriology
May 2006
Vol. 188, #10
Pages 3682-96



Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts

Xiaodong Bai,[1] Jianhua Zhang,[1, 2 a] Adam Ewing,[1 b] Sally A. Miller,[2] Agnes Radek,[3 c] Dimitriy Schevchenko,[3] Kiryl Tsukerman,[3] Theresa Walunas,[3] Alla Lapidus,[3 d] John W. Campbell,[3 e] and Saskia A. Hogenhout[1*]

[1]*Department of Entomology and* [2]*Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center (OARDC), Wooster, OH 44691.*

[3]*Integrated Genomics, Chicago, IL 60612*

Running title: Comparative analysis of two phytoplasma genomes

Keywords: insect-vectored phytopathogen, symbiont, mycoplasma, recombination

Footnotes:

[a]Present address: Potato Research Center, Agriculture and Agri-Food Canada, Fredericton, NB E3B 4Z7, Canada

[b]Present address: GCB Graduate Group, University of Pennsylvania, Philadelphia, PA 19104

[c]Present address: Epicentre Technologies Corp., Madison, WI 53713

[d]Present address: Microbial Genomics, DOE Joint Genome Institute, Walnut Creek, CA 94598

[e]Present address: Laboratory of Genetics, University of Wisconsin, Madison, WI 53706

[*]Corresponding author. Mailing address: Department of Entomology, The Ohio State University, Ohio Agricultural Research and Development Center (OARDC), 1680 Madison Avenue, Wooster, OH 44691. Tel: 330-263-3730. Fax: 330-263-3686. Email: hogenhout.1@osu.edu

Phytoplasmas (*Candidatus* Phytoplasma, Class Mollicutes) cause disease in hundreds of economically important plants, and are obligately transmitted by sap-feeding insects of the order Hemiptera, mainly leafhoppers and psyllids. The 706,569-bp chromosome and four plasmids of aster yellows phytoplasma strain witches' broom (AY-WB) were sequenced and compared to the onion yellows phytoplasma strain M (OY-M) genome. The phytoplasmas have small repeat-rich genomes. The repeated DNAs are organized into large clusters, potential mobile units (PMUs), which contain *tra5* insertion sequences (ISs), and specialized sigma factors and membrane proteins. So far, PMUs are unique to phytoplasmas. Compared to mycoplasmas, phytoplasmas lack several recombination and DNA modification functions, and therefore phytoplasmas probably use different mechanisms of recombination, likely involving PMUs, for the creation of variability, allowing phytoplasmas to adjust to the diverse environments of plants and insects. The irregular GC skews and presence of ISs and large repeated sequences in the AY-WB and OY-M genomes are indicative of high genomic plasticity. Nevertheless, segments of ~250 kb, located between genes *lplA* and *glnQ* are syntenic between the two phytoplasmas, contain the majority of the metabolic genes and no ISs. AY-WB is further along in the reductive evolution process than OY-M. The AY-WB genome is ~154 kb smaller than the OY-M genome, primarily as a result of fewer multicopy sequences, including PMUs. Further, AY-WB lacks genes that are truncated and are part of incomplete pathways in OY-M. This is the first comparative phytoplasma genome analysis and report of the existence of PMUs in phytoplasma genomes.

Phytoplasmas cause disease in over 200 economically important plants, and are obligately transmitted by phloem-feeding insects of the order Hemiptera, mainly leafhoppers and psyllids. They are unique bacteria as they can efficiently invade cells of insects and plants, organisms belonging to two kingdoms. Phytoplasmas are members the Class *Mollicutes*. Mollicutes are soft-skinned (*mollis* = soft, and *cutis* = skin, in Latin) bacteria due to lack of an outer cell wall, and usually have a small genome size, a low (G + C) content, a small number of rRNA operons, few tRNA genes, and limited metabolic activities (15). Mollicutes represent a branch of the phylogenetic tree of the Gram-positive eubacteria, and are most related to the low GC Gram positive bacteria such as *Bacillus*, *Clostridium* and *Streptococcus* spp. (86, 88).

The phylogenetic tree of mollicutes is composed of two major clades that diverged early in evolution (46). One clade contains the orders Acholeplasmatales and Anaeroplasmatales (AAA clade mollicutes), and the other the orders Mycoplasmatales and Entomoplasmatales (SEM clade mollicutes). Phytoplasmas, formerly known as mycoplasma-like organisms of plants, form a monophyletic group in the order Acholeplasmatales (46), and were recently assigned to a novel genus *Candidatus* (*Ca.*) Phytoplasma (84). Approximately 20 phytoplasma phylogenetic groups have been proposed based on 16S rRNA gene sequences, and new branches are continuously being discovered (61, 77). Members of the order Acholeplasmatales are in several ways distinct from other mollicutes. For instance, whereas most mollicutes use UGA as a tryptophan codon instead of a stop codon, a feature they share with mitochondria, the acholeplasmas and phytoplasmas retained UGA as a stopcodon (72).

Mollicutes have been extensively studied because of their economical importance. They are disease agents and obligate inhabitants of humans, mammals, reptiles, fish, arthropods and plants. Phytoplasmas are generally associated with arthropods and plants, whereas mycoplasmas

(Entomoplasmatales and Mycoplasmatales) and ureaplasmas (Mycoplasmatales) are human and animal pathogens causing infections of the respiratory and urogenital tracts, eyes, alimentary canals, glands and joints of humans and animals. Interestingly, three spiroplasmas, *Spiroplasma kunkelii*, *S. citri* and *S. phoeniceum*, are also insect-transmitted plant pathogens, but belong to the order Entomoplasmatales (30), and hence are distantly related to the phytoplasmas. Dual phytoplasma and spiroplasma infections of insects and plants occur frequently (36).

Several mycoplasmas, ureaplasmas, spiroplasmas and acholeplasmas have been cultured outside their hosts in artificial culture medium. Culture media are complex, likely because mollicutes suffered extensive gene losses and, consequently lack genes of basic metabolic pathways. However, so far, phytoplasmas have not been cultured in cell-free medium indicating that phytoplasmas have a different metabolism and probably more reduced genomes than other mollicutes.

The aster yellows phytoplasma strain witches' broom (AY-WB) strain (*Ca.* Phytoplasma asteris; class Mollicutes) generally spreads systemically in lettuce (*Lactuca sativa* L.) and China aster (*Callistephus chinensis* Nees) inducing a variety of symptoms, including vein clearing, yellowing, stunting, witches'-broom, pigment loss or sterility of flowers and necrosis (91). The extreme malformations of plants suggest that phytoplasmas interfere with plant hormone metabolism (46). AY-WB also spreads systemically in *Arabidopsis thaliana* and *Nicotiana benthamiana* inducing yellowing, stunting, and witches'-broom in both (Bai, Correa, and Hogenhout, unpublished results). AY-WB was classified into the 16SrI-A subgroup of *Ca.* Phytoplasma asteris, based on the restriction fragment length polymorphism (RFLP) banding pattern of a 1.2-kb 16S rDNA polymerase chain reaction (PCR) fragment (91). In contrast, OY-M, the only other phytoplasma for which a complete genome sequence is available (66), belongs

to the 16SrI-B subgroup (46). *Ca.* Phytoplasma asteris, previously known as aster yellows phytoplasma (AYP) or group I phytoplasma (47), is the largest of the phytoplasmas and associates with more than 100 economically important diseases worldwide (46, 57). Plant hosts include broad-leaf, herbaceous plants and several woody fruit crops (57).

AY-WB is transmitted by the polyphagous leafhopper *Macrosteles quadrilineatus* (Forbes). Phytoplasma interactions with insects are complex and involve intra- and extracellular replication in gut and salivary glands epithelial and muscle tissues, and other organs and tissues. Whereas there is evidence that some phytoplasmas are vertically transmitted to the progeny of their insect vectors (33), the predominant means of survival of phytoplasmas is through transmission between insects and plants. They appear to manipulate their insect and plant hosts to enhance their own transmission efficiency. For example AYPs increase fecundity and longevity of their insect vector *M. quadrilineatus* (11).

Because of their small genomes and economic importance, mollicutes have been targeted for genome sequencing projects for some time. *Mycoplasma genitalium* was the second bacterium sequenced to completion because of its minimal gene complement for a cultivable organism (29). Thus far, genomes of nine SEM clade mollicutes and one AAA clade mollicute (OY-M phytoplasma) (68) have been fully sequenced. Here, we report the full sequence of the small genome of AY-WB. Comparative genome analysis revealed the presence of 14 to 23 % repetitive DNA organized in putative mobile units (PMUs) in the phytoplasma genomes, and differences in standard metabolic and non-metabolic pathways between phytoplasmas and SEM clade mollicutes.

MATERIAL AND METHODS

DNA isolation. The AY-WB strain was collected from diseased lettuce plants in Celeryville, Ohio (41.00ºN, 82.45ºW) in 1998 (91). AY-WB was isolated from lettuce plants about two weeks after symptom appearance. The stems of lettuce plants were cut at several places with a sharp razor blade, and phloem sap oozing from the cut area was collected. On average, 1.6 ml sap was collected from each symptomatic lettuce plant. For preparation of gel plugs, 200 µl sap was immediately mixed with 800 µl pre-cooled 30% glucose-1X TE (pH 8.0) buffer, followed by centrifugation at 16,000 × g for 20 min at 4°C. The pellet was mixed with 80 µl 1% pre-melted low melting agarose (45°C) in 0.5X TBE (pH 8.0) and incubated at 4°C. Solidified plugs were subjected to proteinase K digestion at 50°C for 48 h, and then rinsed with 1X TE buffer (pH 8.0) three times before subjection to pulsed field gel electrophoresis (PFGE). PFGE was conducted in a 1% agarose gel with a running time of 18 h, 60–120 sec switch time ramp, voltage of 6V/cm and an included angle of 120° (CHEF-DR III, Bio-Rad, Hercules, California, U.S.A.). The AY-WB chromosome produced a single band of ~700 kb in the PFGE gel. The identity of the band was confirmed by Southern blot hybridizations and PCR using phytoplasma-specific probes and primers, respectively. The 700-kb fragment was excised from the gel, and the gel blocks were placed directly into the Elutrap (Schleicher & Schuell) collection chamber for elution of DNA at 106 V at 4 ˚C for 15 h. DNA was ethanol-precipitated using standard procedures and resuspended in deionized distilled water. The concentration of the purified genomic DNA was assessed using a PicoGreen kit (Molecular Probes).

Sequencing strategy. The shotgun library was constructed at Integrated Genomics Inc. (IG). Five microgram DNA was sheared using a computer-controlled shearing device (GeneMachines,

San Carlos, California, U.S.A.) to produce DNA fragments of 2 kb on average. Sheared DNA was loaded onto 0.7% agarose gels and DNA fractions corresponding to 2-2.5 kb were extracted from the agarose gel. Single-stranded ends of the DNA were removed by T4 polymerase and then filled in with Klenow fragment. Size-selected 2-2.5 kb DNA fragments were cloned into the pGEM-3Z vector (Promega, Madison, WI), introduced into *Escherichia coli* DH10B, and sequenced with the DYEnamic ™ ET Dye Terminator Kit (Amersham Biosciences, Piscataway, NJ). Sequence quality assessment and subsequent assembly were performed with the Phred/Cross_match/Phrap package (25, 26) and PGA (Paracel Genome Assembler). Sequencing and physical gaps in the assembly were closed by multiplex PCR (83) and primer walking.

Annotation. The sequence data of AY-WB were submitted to the IG database and software suite, ERGO, for sequence annotation. CRITICA (7), Glimmer2 (21) and IG-proprietary tools were used for open reading frame (ORF) identification. ORF function annotation was conducted by a number of IG-proprietary algorithms that automatically predict the function of ORFs based on comparative analysis with orthologues clusters in ERGO. In addition, the predicted proteins were searched, using the BLAST algorithm (6), against a non-redundant (nr) database at the National Center for Biotechnology Information (NCBI). Protein functional domains were analyzed by searching against the NCBI conserved domain database (55) and the Pfam database (10). The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used for the reconstruction of the metabolic pathways. The assignment of Enzyme Commission (EC) number was according to the BRENDA database (78).

Database submission. Sequences of the AY-WB genome have been deposited at GenBank under accession numbers CP000061 (chromosome), CP000062 (plasmids AYWB-pI), CP000063 (plasmid AYWB-pII), CP000064 (plasmid AYWB-pIII) and CP000065 (AYWB-pIV). More

detailed information on the AY-WB genome is available on our website http://www.oardc.ohio-state.edu/phytoplasma

RESULTS

General genomic features. The AY-WB genome is composed of one circular chromosome of 706,569 bp (Fig 1A), and contains two ribosomal RNA (rRNA) operons, 31 transfer RNA genes, and 671 predicted ORFs (Table 1). UGA was used as stopcodon for prediction of the ORFs. This is consistent with other reports showing that acholeplasmas and phytoplasmas retained UGA as a stopcodon, unlike SEM branch mollicutes, which use UGA as a tryptophan codon instead of a stop codon (72). This is also in agreement with annotations conducted for OY-M (68). Our results were not in agreement with a report stating that UGA should be considered as a tryptophan codon in phytoplasmas, as in mycoplasmas (Melamed et al. J Bacteriol. 2003;185:6513-21). The average guanine (G) and cytosine (C) content of the AY-WB chromosome is 27%. The genome has an irregular GC skew pattern different from most prokaryotic genomes, which usually consist of two major shifts near the origin of replication and terminus of replication (31). Irregular GC skew patterns were also found in the genomes of some other bacteria, such as *Wolbachia pipientis* (89) and *M. mycoides* (87). Because the location of the origin of replication (*oriC*) was not clear, the first nucleotide of the *dnaA* gene was assigned basepair (bp) 1. However, the *oriC* is most likely located upstream of the *dnaA* as predicted by the Oriloc software (28) and by the opposite direction of ORFs surrounding the putative *oriC* (Fig. 1A) (31).

In addition to the chromosome, four small circular plasmids were identified (Fig. 1B; Table 2). This was surprising, because the DNA isolation procedure should not allow the isolation of small DNAs. One explanation of this discrepancy is that the plasmids are present at high copy numbers in the phytoplasma cell. As a consequence some plasmid DNA was co-purified from the PFGE gel along with the AY-WB chromosomal DNA. The plasmids contain 22 putative ORFs, and their average GC contents ranged from 21.8% to 25.6%. Each plasmid has genes for a replication initiation protein (Rep) and a single-stranded DNA binding protein (SSB) that are involved in rolling-circle amplification (40), whereas the functions of the other genes are not known. However, most of the plasmid genes were predicted to encode secreted or membrane proteins (Fig. 1B), and except ORF pIII02 of AYWB-pIII and pIV06 of AYWB-pIV, all are similar to OY-M phytoplasma sequences (Table 2). It is striking that, whereas the plasmids encode different Rep proteins, they contain paralogous genes in similar order (Fig. 1B). Two AY-WB plasmids (AYWB-pI and AYWB-pIII) contain *repA* genes similar to geminiviruses *repA*, whereas the *rep* genes of the other two plasmids (AYWB-pII and AYWB-pIV) were unique to AY-WB and OY phytoplasmas.

The AY-WB plasmids seem prone to mutation. First, ORFs pIII04 and pIII05 of AYWB-pIII are respectively similar to the 5' and 3' portions of paralogous genes on the other three plasmids, suggesting that a mutation to a stop codon produced two ORFs in AYWB-pIII. Further, the sequence between pII03 and *ssb* of AYWB-pII is similar to genes pI04, pIII06 and pIV04 of the other three plasmids, but was not annotated as an ORF because of the presence of a premature stop codon. In addition, plasmids apparently recombine with the chromosome, as the latter contains three truncated ORFs similar to the geminivirus-like *repA* plasmid genes and one truncated copy similar to the other *rep* gene (Fig 1C).

Repetitive and mobile DNA in the AY-WB genome. The AY-WB genome contains long repeating units of DNA. Of the 671 predicted ORFs of AY-WB, 191 (28%) covering 97,374 bp (13.8%) of the AY-WB chromosome are present as multiple copies (Fig. 2A). Of these 191 ORFs, 134 (20%) covering 71,979 bp (10.2%) of the chromosome are organized as clusters, consisting of genes encoding transposases (*tra5*), DNA primase (*dnaG*), DNA helicase (*dnaB*), thymidylate kinase (*tmk*), Zn-dependent protease (*hflB*), DNA-binding protein HU (*himA*), single stranded DNA binding protein (*ssb*), a specialized sigma factor (*sigF*), and a number of other genes with unknown function (Fig. 3). Many of these hypothetical proteins are predicted to target the phytoplasma membrane (Figs. 1 and 3; Table 3) and therefore are likely involved in AY-WB interaction with plant and insect hosts.

The phytoplasma *tra5* ISs belong to group IS*150*, family IS*3* (48, 53). The presence of *tra5* insertion sequences (ISs) and other genes involved in recombination and repair, such as *himA*, suggest that these cluster are mobile elements, and hence were named potential mobile units (PMUs). PMU1 is flanked by a complete *tra5* IS on one side and a truncated *tra5* IS at the other side, and inverted repeats of 327 bp (Fig. 3A). Sequences highly similar to the PMU1 inverted repeats were also found adjacent to the *tra5* ISs of the other three PMUs (Fig. 3A). Another striking observation is that all PMUs contain copies of *dnaG*, *dnaB*, *ssb* and *tmk* that are involved in DNA replication, suggesting that the PMUs may transpose in a replicative fashion.

The AY-WB genome also contained several clusters that look like derivatives of PMUs as they contained truncated versions of PMU ORFs with similar gene orders as PMUs. It is likely that these PMU-like clusters are in the process of being eliminated. Based on the positions of the *tra5* insertion sequences, the PMUs or PMU-like ORF clusters are present in at least seven

locations in the AY-WB chromosome (Fig. 1A). At three locations in the AY-WB genome, PMUs are located adjacent to each other. The largest PMU-rich region of the AY-WB chromosome is ~ 75,000 bp (Fig. 1A), including PMUI and PMUII (Fig. 3A).

Not all *dnaG*, *dnaB*, *tmk*, *hflB*, *himA* and *ssb* genes are part of PMUs or PMU-like clusters. As discussed above, several *ssb* genes are located on plasmids or in plasmid-derived sequences within the chromosome (Fig. 1B,C). The AY-WB chromosome also contains single copies of *dnaG*, *dnaB*, *tmk*, *himA* and *hflB* homologs, which are clearly different in sequence from the PMU genes. Further, AY-WB contained several multicopy sequences that are not part of PMUs, including one complete and several truncated copies of *uvrD* and *dam*.

Comparative genome analysis of phytoplasmas. The AY-WB chromosome is 154,062 kb smaller than that of OY-M, and AY-WB has 83 fewer ORFs than OY-M (Table 1). This difference in genome size is the result of a lower number of multicopy genes in AY-WB compared to OY-M (Fig. 2A). OY-M multicopy genes are also organized in PMUs. The AY-WB genome contains 97,374 bp (13.8%, 191 ORFs) multicopy sequences compared to 195,035 bp (22.7 %, 268 ORFs) for OY-M, and the majority are clustered in PMUs with 71,979 bp (10.2%, 134 ORFs) for AY-WB and 121,226 bp (14.1%, 175 ORFs) for OY-M. Thus, compared to OY-M, the 154,062-kb smaller genome of AY-WB is due to 97,661 kb fewer multicopy genes. The percentages of non-coding DNA are similar between AY-WB and OY-M, but because the OY-M genome is larger, OY-M non-coding DNA absorbs an additional 55,728-bp genome size difference between AY-WB and OY-M (Fig. 2A). As expected based on these observations, the numbers of single copy ORFs are similar between the phytoplasmas with 432,553 bp (482 ORFs, 61.2%) for AY-WB and 433,226 bp (486 ORFs, 50.3%) for OY-M (Fig. 2A).

Alignment of the AY-WB and OY-M genomes has an X-shaped pattern illustrating synteny of the majority of AY-WB and OY-M sequences, but inverse orientation of large genome segments (Fig. 2C). In both AY-WB and OY-M, the largest aligned region is ~250 kb and starts with gene *lplA* at 423,992 bp in AY-WB and 354,087 bp in OY-M and ends with *glnQ* at 660,824 bp in AY-WB and 103,752 bp in OY-M (arrowheads in Fig. 2C). This region is upstream of the putative *oriC* in AY-WB, but downstream of the putative *oriC* in OY-M. In both AY-WB and OY-M, these ~250 kb regions contain the majority of the metabolic genes, and do not contain *tra5* insertion sequences (Fig. 1).

The PMUs tend to congregate as evidenced by the groups of ISs, and are frequently located on opposite strands as can be noticed by the correlation of GC skew inflection points and the boundaries of sense-antisense regions, and *tra5* insertion sequences in the AY-WB chromosome (Fig. 1A). The alignment of the AY-WB and OY-M chromosomes revealed that PMUs or PMU-like sequences at six locations in the AY-WB chromosome are also present at the same locations in the OY-M chromosome. However, at three locations the sequences in AY-WB or OY-M have undergone excessive deletion and mutation events. PMU sequences at one location in the AY-WB chromosome and four locations in the OY-M chromosome are unique to each of the phytoplasmas. Like AY-WB, the OY-M genome contains several genes that are not part of PMUs, including two full-length and several truncated copies of *dam*, and three full-length and several truncated copies of *uvrD*. Our observations are consistent with those of others as Oshima et al. (2004) reported that the OY-M genome contains multiple copies of *uvrD, hflB, tmk, dam* and *ssb* constituting 18% of the total genes.

Besides the PMUs and other multicopy sequences, other differences between AY-WB and OY-M were found. Strikingly, AY-WB lacks most sequences that are truncated in OY-M (Fig

2B), including *hsdR* and *hsdM* of the type I restriction-modification system, three adjacent fragments with similarities to *recA*, and two adjacent sequences of the *sucP* gene for sucrose phosphorylase (EC: 2.4.1.7). AY-WB also lacks genes that are part of incomplete pathways in OY-M, including *rfaG* (EC: 2.4.1.157) of the glycerolipid metabolism pathway, and *pdxK* (EC: 2.7.1.35) of the vitamin B6 pathway. Finally, whereas AY-WB lacks *folC* (EC: 6.3.2.17) and has truncated versions of *folK* (EC: 2.7.6.3) and *folP* (EC: 2.5.1.15), OY-M has full-length copies of these genes that belong to the folate biosynthesis pathway. Only a few AY-WB ORFs with functional annotations were absent from OY-M (Fig.2B). These include *cbiQ* and *evbH* of the cobalt and multidrug ATP-binding cassette (ABC) transporter systems, respectively (Table 4). However, OY-M has chromosome fragments with similarities to *cbiQ* and *evbH*, but ORFs were not assigned. Except for these sequences, a high degree of gene content conservation was observed between the genomes of AY-WB and OY-M, including major metabolic pathways, and ABC and P-type ATPase transporters (68) (Tables 4 and 5).

Comparative genomics of phytoplasmas and other mollicutes. To determine to what extent phytoplasma genomes differ from the distantly related SEM clade mollicutes, ORF sequences of the AY-WB and OY-M phytoplasmas were compared to those of nine *Mycoplasma* and *Ureaplasma* spp. (blastp, E value < 10^{-5}). More than half of the phytoplasma ORFs had similarities to those of SEM clade mollicutes, and AY-WB and OY-M had an equal number of phytoplasma unique ORFs (318 ORFs) (Fig. 2D). Relative to OY-M, AY-WB contained fewer ORFs that were present in several but not all SEM branch mollicutes (146 ORFs for AY-WB vs. 214 ORFs for OY-M; Fig. 2D). The ~250 kb segment between genes *lplA* and *glnQ* that is syntenic between the AY-WB and OY-M phytoplasmas (Fig 2C) contained the majority of the ORFs conserved among mollicutes (blue patches in ring 5 of Fig. 1), while the less syntenic

region (first 400 kb of the AY-WB genome; Fig. 2C) are repeat-rich (IS elements ring 4 Fig. 1; Fig. 2C) and are more enriched with phytoplasma-specific ORFs (red patches of ring 5 Fig. 1).

Of the 318 ORFs that are unique for phytoplasmas in the class Mollicutes, 40 had functional annotations and were closely examined (Table 6), since these may be part of metabolic pathways absent from SEM branch mollicutes. These 40 ORFs include *sfcA* for NAD-specific malic enzyme (EC: 1.1.1.38) and two copies of the malate/citrate-sodium symporter genes *citS*. Phytoplasmas have a maltose ABC transporter system, including a maltose binding protein (MalE) (Table 4) and several other transporters that are not present in the SEM clade mollicutes (Table 6). These include several components of the *art* and *gln* ABC transporter systems that might be important for import of glutamine and arginine, respectively, and several solute-binding proteins, including ArtI predicted to bind arginine (35), the dipeptide binding protein and D-aminopeptidase DppA (17), and NlpA lipoprotein (90) for which the gene is located between methionine ABC transporter genes and hence may produce a methionine binding protein (Table 4). Phytoplasmas also have *mntB* and *znuA* of the manganese (Mn) and zink (Zn) ABC transporter system (13) (Table 6). All the solute-binding proteins were predicted to have signal peptides (SignalP v3.0) (12) and are likely extracellular lipoproteins (34). Two ABC transporters have adjacent genes for thermostable carboxypeptidase 1 (EC: 3.4.17.19) and oligoendopeptidase F (EC: 3.4.24.-) that can process imported peptides and were not present in the genomes of SEM branch mollicutes (Table 6). Finally, three AY-WB genes were annotated as *norM* that encodes a Na^+-driven multidrug efflux pump. One *norM* gene had similarity to genes of SEM mollicutes, whereas the other two did not. These two are located adjacent to each other and are transcribed in opposite directions in both the AY-WB and OY-M genomes.

Other genes present in AY-WB and OY-M, but absent from SEM-branch mollicutes are *pssA* and *psd* (Table 6) of the phosphatidylethanolamine (PE) pathway (58). Further, mycoplasmas lack *pcnB* encoding poly(A) polymerase (EC: 2.7.7.19) and *pnp* encoding polyribonucleotide nucleotidyltransferase (PNPase; EC: 2.7.7.8). Both are involved in the regulation of mRNA stability. Interestingly, the *pnp* gene is present in the genome of *S. kunkelii* (8), which is also an insect-transmitted plant pathogenic mollicute. PNPase may be involved in the persistent infection of insects and/or adaptation to diverse hosts and habitats of phytoplasmas and spiroplasmas (8). The adjoining phytoplasma genes *pmbA* and *tldD* were not identified in SEM branch mollicutes either. PmbA and TldD regulate DNA gyrase function and are involved in protein maturation (3, 62, 75).

Compared to other mollicutes, phytoplasmas lack several essential transporters and pathways. AY-WB and OY-M lack phosphoenolpyruvate:sugar phosphotransferase (PTS) systems for import of sugars essential for glycolysis. AY-WB and OY-M also lack F-type ATPases. This is in contrast to mycoplasmas and ureaplasmas that have ATPase complexes, including the A, B and C subunits for the transmembrane channel, and the five-subunit (α, β, gamma, delta, epsilon) catalytic core for ATP synthesis, and can generate a transmembrane potential with resultant ATP synthesis (72). However, phytoplasmas have five genes encoding P-type ATPases (Table 5) that may generate electro-chemical gradients over the membrane.

Phytoplasmas have fewer genes in the standard recombination pathway and SOS response in comparison to SEM branch mollicutes. All mollicutes sequenced so far lack *recB, recC, recD, recG* and *ruvC* of the recombination pathway, and *recN, recO, recQ* and *recR* of the SOS response, although some mycoplasmas carry *recR* and *recO*. Thus, SEM branch mollicutes have *recA, recU, Ssb, polA, gyrA, gyrB, ruvA* and *ruvB*, a rudimentary set of genes that permit

homologous recombination. Of these, phytoplasmas do not have *recA*, *ruvA* and *ruvB*. Hence, phytoplasmas have a deficient homologous recombination machinery.

AY-WB virulence. The AY-WB genome was analyzed for similarities to known bacterial virulence factors. Several putative hemolysins of AY-WB were identified based on annotation. These include a protein annotated as HlyC, a putative hemolysin III. This protein belongs to integral membrane protein family (Pfam domain # PF03006), which includes a protein with hemolytic activity from *Bacillus cereus*. However, other proteins in this family play a role in lipid and phosphate metabolic pathways. Another putative hemolysin-related protein of AY-WB was annotated as TlyC, a putative hemolysin-related protein, which carries resemblance to Cluster of Orthologous Group (COG) 1253 of hemolysins and related proteins containing CBS domains. Indeed, AY-WB TlyC contains a CBS domain (Pfam domain # PF00571). However, the AY-WB TlyC protein has a N-terminal transmembrane region (Pfam domain # PF01595) not found in TlyC proteins, and a C-terminal domain that is present in the C-terminus of Na+/H+ antiporters, including CorC involved in magnesium and cobalt efflux (Pfam domain # PF03471). Thus, it is not clear whether HlyIII and TlyC of AY-WB are hemolysins.

Two AY-WB proteins, AYWB084 and AYWB352, are similar to the *Legionella pneumophila* virulence factor IcmE (E-values $5e^{-21}$ and $5e^{-05}$, respectively), which is part of the type IVB secretion system apparatus that translocates bacterial proteins into host cells (79). Proteins with similarities to IcmE were also identified in the OY-M genome (68). IcmE has sequence similarity to plasmid genes involved in conjugation (79). In both AY-WB and OY-M the majority of the *icmE*-like sequences were located upstream of the ATP-dependent helicase gene *uvrD*. UvrD belongs to the Rep family helicases and catalyzes ATP-dependent mediated unwinding of double-stranded DNA into single-stranded DNA, and has a role in the recF

recombination pathway, methyl-directed mismatch repair, UvrABC-mediated nucleotide excision repair and replication (32, 59). Similarly to the other repeated sequences, the OY phytoplasma genome contains multiple copies of *icmE*-like sequences and full-length *uvrD*, whereas the AY-WB contains only one full-length *icmE*-like sequence and *uvrD* and multiple truncated copies of these sequences. Further research should reveal whether the *icmE*-like sequences of phytoplasmas mediate conjugation or are somehow involved in the recombination pathway. No other similarities of phytoplasma sequences to type III and type IV secretion systems were observed. This may not be surprising as translocation of virulence factors via type III and IV secretion systems is more specific for Gram-negative bacteria.

AY-WB and OY-M share the genes of the protein export and targeting components of the sec-dependent pathway, including *secA, secY, yidC, ffh, ftsY, dnaJ, dnaK, grpE, groES*, and *groEL*, and, like SEM branch mollicutes, lack several subunits and the signal peptidases of the protein maturation component, including *secB, secG, secF, secE, secD* and signal peptidase Spase I (72). Despite the absence of several components, OY-M phytoplasma has a functional sec-dependent protein translocation system (38). It is possible that some of the many hypothetical proteins have peptidase activity. This confirms other findings (9, 39) that phytoplasmas have a functional protein sec-dependent protein translocation system and that the N-terminal signal peptides of proteins are cleaved. Since the closest walled relatives of phytoplasmas are *Clostridium*, *Bacillus* and *Streptococcus* spp. (Phylum Firmicutes), it is possible that, similarly to *Streptococcus pyogenes* (76), phytoplasmas secrete virulence-related proteins via the sec-dependent pathway.

Both phytoplasma genomes contain several ATP-binding cassette (ABC) transporters (Table 4). ABC transporters import peptides, amino acids and nutrients into the cell, and can be

virulence factors as well as they can deplete the host from essential nutrients, and secrete toxins and antimicrobial compounds such as hemolysins (19). Further, solute-binding proteins of ABC transporters are usually secreted lipoproteins that bind substrate external to the cell and deliver the substrate to the ABC transporters, and may also be involved in adherence to cell surfaces (4). For instance, the ABC transporter related solute-binding protein Sc76 of *Spiroplasma citri* was shown to be involved in penetration of or multiplication in the salivary gland (14). The AY-WB genome contains genes for five solute-binding proteins with specific solute-binding activities (Table 4). All five solute-binding proteins have N-terminal cleavable signal peptide sequences as predicted with the SignalP v3 software (12), and therefore are secreted via the sec-dependent pathway. Hence, these five solute-binding proteins are putative virulence factors of phytoplasmas.

DISCUSSION

It is intriguing that phytoplasmas have small genomes, which lack many standard metabolic functions, but are repeat rich. The repeated DNAs are mostly multicopy genes organized in potential membrane units (PMUs). Thus, phytoplasmas are different from other bacterial endosymbionts of insects, e.g. *Buchnera* and *Blochmannia* spp., which also have small genomes lacking many standard metabolic functions, but have low levels of repeated DNAs (1, 82). On the other hand, the majority of the mollicutes have repeat-rich genomes. All mollicutes are under pressure for genome minimization, and the presence of numerous repeats is therefore highly significant (74). Indeed, for several mycoplasmas it has been shown that repeats engage in recombination events resulting in changes of mosaics of antigenic structures at cell surfaces, essential for evasion of the host immune system and for adaptation to new environments (74).

Thus, similarly to mycoplasmas, the repeated DNAs of phytoplasmas probably allow adaptations to different environments. Adaptation is particularly important for phytoplasmas, as their host environments are extremely variable, including the intracellular environments of phloem tissues of plants, and guts and salivary glands and other organs and tissues of insect hosts. Also, phytoplasmas have a broad host range. AY-WB alone can infect China aster, lettuce, tomato, *Nicotiana benthamiana* and *Arabidopsis thaliana*. Phytoplasma genomes are in several aspects different from mycoplasma genomes. Firstly, phytoplasmas do not have *recA*, *ruvA* and *ruvB*, and hence appear to lack a functional recombination system. Secondly, thus far, the organization of repeated DNAs in PMUs is unique to phytoplasmas among the mollicutes.

PMUs. The PMUs contain *tra5* insertion sequences (ISs), which belong to group IS*150*, family IS*3* (48, 53). IS3 type mobile units are found in a number of other mollicutes, for example IS*1138* in *M. pulmonis*, IS*1221* in *M. hyorhinis* and *M. hyopneumoniae*, IS*1297* in *M. mycoides* subsp. *mycoides*, ISMi*1* in *M. incognitos*, and one IS*3* element in the spiroplasma virus DNA SPV1-C74 sequence of *S. citri* (53, Melcher et al., Microbial & Comp. Genomics 4:29). All belong to the IS*150* subgroup, and some of these elements have been demonstrated to undergo autonomous transposition (Bhugra and Dybvig, 1993. Mol. Microbiol. 7: 577-584).

PMU1 of AY-WB is the longest and appears most complete, and has several striking features characteristic of composite transposons (Fig. 3A). First, the right and left borders of PMU1 contain long (327 bp) inverted repeats (IRs). Further, whereas the ORF to the right is a truncated *tra5* sequence, the *tra5* sequence at the left can produce a full-length ORFAB fused-frame transposase (53). IS*150* can generate circles by joining IRs upon production of the fused-frame transposase (81), and particularly composite transposons that carry single inverted repeats at the left and right borders form stable circles (43). PMU1 also carries a gene for DNA protein HU

(*himA*), which is a non-specific binder of DNA but prefers binding to bent, kinked or altered DNA sequences (27) and has a role in recombination through the joining of distant recombination sites (5). Thus, with the help of transposase and DNA protein HU, the IRs could join to form a circle and induce transposition of PMU1. It is striking that all the genes on PMUs are oriented in the same direction with *sigF*, encoding a specialized transcription factor, as the first gene and located downstream of the inverted repeat. In IS*3* family members, the adjoined IRs, which are formed on circularization, create a strong hybrid promoter that drives high levels of transposase expression (53). Hence, it is possible the adjoined 327-bp repeats upon circulation of PMU1 creates a strong promoter that drives the expression of, at least part, of the PMU genes.

The AY-WB and OY-M genomes also contain evidence that at least some PMUs transpose in a replicative fashion. Firstly, there are multiple copies of PMUs and PMU-like clusters. Secondly, the PMUs contain full-length *dnaB, dnaG,* and *ssb* genes that are involved in DNA replication. DnaB initiates DNA replication (16). It moves along the lagging strand and unwinds the DNA helix for the propagating fork, and attracts DnaG for lagging strand synthesis (85). SSB plays an essential role in DNA replication by stabilizing single-stranded DNA (51). Most PMUs also contain a *tmk* gene encoding thymidylate kinase that synthesizes dTDP from dTMP for DNA synthesis. Similarly to AY-WB, the OY-M phytoplasma genome contains at least two *tmk* homologs, *tmk-a* and *tmk-b*, with *tmk-a* being present as multiple copies (60). We revealed that the *tmk-a* genes are part of PMUs. However TMK-b but not TMK-a was shown to have thymidylate kinase activity (60). Hence, the function of TMK-a is not yet clear.

Several sigma factor genes were identified in the AY-WB genome. These are *rpoD* that encodes the standard 465 amino acid sigma70 protein and is present as a single copy on the AY-WB chromosome, and multiple copies of *sigF* that are located on PMUs or PMU-like gene

clusters and have deduced protein of ~200 amino acids in lengths. PMU3 contains sequence with similarity to *sigF* immediately upstream of the *ssb* gene, but because of the presence of a premature stopcodon, this sequence was not predicted to be an ORFs. The OY-M genome also has multiple copies of *sigF* that are part of PMUs. The N-terminal 100 amino acids of the SigF proteins have region 2 domains (pfam04542) containing both the -10 promoter recognition helix and the primary core RNA polymerase binding determinant. However, the C-terminal 100 amino acids of the SigF proteins do not have similarities to other proteins or domains, including the region 4 domains (pfam04545) containing the -35 promoter-binding element. AY-WB SigF proteins showed greatest similarities (E-value 10^{-6}) to the stress-response sigma factor (sigma(H)) of *Streptococcus coelicolor* (45) and the flagellar biosynthesis sigma factor FliA of *Pseudomonas putida* (41). Expression of SigF and other PMU genes might occur under specific environmental conditions.

Since PMUs contain several genes predicted to encode membrane-targeted sequences, one would expect that expression of PMU genes would result in a change of the phytoplasma membrane surface. In this regard, it is intriguing that the PMUs contain *hflB* (or *FtsH*) genes encoding membrane-associated ATP-dependent Zn proteases of ~700 amino acids. These proteins are conserved among bacteria, and are involved in membrane-associated processes such as protein secretion (22) and membrane protein assembly (2), as well as adaptations to nutritional conditions and osmotic stress (22,52).

Genomic plasticity. The irregular GC skews and presence of large repeated sequences (PMUs) in the AY-WB and OY-M genomes are indicative of high genomic plasticity. The correlation between an irregular GC skew and presence of ISs in mollicute genomes is quite striking. For instance, *M. mycoides* has an irregular GC skew and 13% of the genome size

consists of ISs (87), whereas *M. mobile* has a regular GC skew and no ISs (37). It should be noted, however, that although AY-WB doesn't have a significant GC skew, it may have another kind of significant skew or excess, including AT skew, and purine excess or keto excess (Song et al., 2003. BMC Genomics).

Phytoplasma genomic plasticity is also evidenced by the differences in genome sizes and compositions between members of *Ca*. Phytoplasma asteris, ranging from 660 to 1,130 kb, and consisting of several fragments of 500 kb and larger (56, personal observation). Since PMUs can form large clusters that may locate in different sections of the chromosome, it is likely they are also capable of splitting a single chromosome into two smaller chromosomes. Further, results reported herein show that AY-WB and OY-M differ ~154 kb in genome size, mainly because of a difference in PMUs and other multicopy sequences (Fig. 2A).

Despite the phytoplasma genome plasticity, the majority of the AY-WB and OY-M genomes are syntenic (Fig. 3C). Scatterplots of conserved sequences between the AY-WB and OY-M genomes shows an X-shaped pattern with symmetry around the tentative *oriC*, and two other locations at approximate opposite ends of the *oriC* (Fig. 3C). This X-shaped pattern or X alignment is common in genome comparisons of closely related bacterial species, and is most likely due to the occurrence of large inversions that rotate around the oriC and terminus of replication (24). The breakpoints of the inversions between the AY-WB and OY-M genomes are, as expected, at PMU-like regions, and repeated *uvrD* sequences.

There are probably two reasons for the good alignment of the AY-WB and OY-M genomes. Firstly, we already observed that the PMUs tend to congregate. This is consistent with findings that IS*150* frequently transpose into target regions resembling their IR (53, 66). Thus, transposition will predominantly affect certain areas of the phytoplasma genomes, and hence the

synteny in the rest of the genome can be maintained. Secondly, because of the absence of *recA*, *ruvA* and *ruvB*, rearrangements between PMUs through homologous recombination are likely to occur at lower frequencies than in genomes with RecA-dependent homologous recombination machineries (70, 71).

Variations in the presence of *recA* are common among insect-associated mollicutes (Melcher and Fletcher, 1999; Eur J Plant Pathol. 105:519). Truncated *recA* genes were found in six *Spiroplasma citri* strains, which like phytoplasmas are insect-transmitted plant pathogens, and five *S. melliferum* strains, which are pathogens of bees (54). In *S. citri* only the first 390 nucleotides at the 5' end of *recA* are present, whereas in *S. melliferum* the full-length *recA* gene is interrupted by a TAA stopcodon. Intriguingly, truncated and full-length RecA polypeptides were observed in a proteomic study of *S. melliferum* (Cordwell et al. 1997, Electrophoresis 18: 1335). These finding suggest that *recA* sequence variation among insect-associated mollicutes is of biological significance. RecA has an important function in mycoplasmas. Deletion of *recA* is lethal for *M. pulmonis* (72). RecA is probably essential for homologous recombination between repeated lipoproteins, and adhesin genes result in a change of mosaic of antigenic structures at the bacterial surface, with subsequent evasion of the host immune response (72, 74). Thus, it seems that phytoplasmas and spiroplasmas can adapt to their hosts with a less efficient homologous recombination system, and loss of RecA function might then be beneficial for increasing genome stability. This is supported by the observations that, like phytoplasmas, spiroplasmas have highly repeat-rich genomes mainly due to phage-derived sequences (72). On the other hand, *M. mycoides*, which also has a repeat-rich genome and is a human pathogen, has a full-length *recA* (42).

Reductive evolution. In general, AY-WB seems further along in the reductive evolution

process than OY-M. Firstly, AY-WB phytoplasma contained fewer PMUs insertions, and the ORFs in AY-WB PMUs are more frequently truncated or deleted. Secondly, AY-WB lacks genes that are truncated in OY-M, including *asnB, hsdR, hsdM, recA* and *sucP*. Thirdly, AY-WB lacks genes of incomplete pathways in OY-M, including *rfaG* of the glycerolipid metabolism pathway and *pdxK* of the vitamin B6 pathways. Further, unlike OY-M, AY-WB does not have *folC,* and OY-M has full-length *folK* and *folP* genes that are truncated in AY-WB. The *folK* and *folP* genes were also identified as pseudogenes in clover phyllody (CPh) phytoplasma (*Ca.* Phytoplasma asteris) (20), suggesting that OY-M may be capable of *de novo* folate synthesis, whereas AY-WB and CPh have to import folate from host cells. Similarly to CPh (20), the *folK* and *folP* sequences of AY-WB and OY-M are flanked by *gcp*, which encodes a glycoprotease, and two ORFs encoding a DegV family protein and a 24-kDa lipoprotein (AYWB245) (20). Hence, the gene organizations of this part of the genome are conserved among *Ca.* Phytoplasma asteris members. Final evidence that AY-WB is further down the reductive evolutionary path is provided by the observation that, relative to OY-M, AY-WB contained fewer ORFs that are shared by several but not all mollicutes (146 ORFs for AY-WB vs. 214 ORFs for OY-M; Fig. 2D).

Plasmids. We identified four plasmids in AY-WB. Plasmids have been detected in a number of other phytoplasmas (50, 65). Each AY-WB plasmid contains two genes involved in rolling circle amplification, and two to six ORFs with unknown function of which several were predicted to target the AY-WB membrane suggesting that the plasmids are involved in AY-WB association with the plant and insect hosts. Indeed, the RepA proteins of OY-M phytoplasmas were detected in infected plants (63), indicating that the plasmid genes are expressed during

infection of the plant. Further, spontaneous OY-M mutants, which lack ORFs on a plasmid and are non-insect transmissible, were isolated (64).

Interestingly, two AY-WB plasmids (AYWB-pI and AYWB-pIII) contain *repA* genes similar to geminivirus *repA*, whereas the *rep* genes of the other plasmids were unique to AY-WB and OY-M phytoplasmas. Geminivirus-like *repA* genes were also identified in OY-M (67), and more distantly related phytoplasmas (50, 73). Like phytoplasmas, geminiviruses are insect-transmitted plant pathogens and have to pass through the gut epithelium, hemolymph and salivary gland cells of the insect vectors before returning to the plant (18). Phytoplasmas and geminiviruses have overlapping plant and insect host ranges. Hence, it is possible that phytoplasmas acquired the *repA* genes from geminiviruses through horizontal exchange. On the other hand, it has been hypothesized that geminiviruses were originated from bacterial plasmids (44). Plasmids with similar *repA* genes are generally incompatible and therefore it is likely that the four plasmids are not present in one AY-WB cell, but represent the plasmid content of the AY-WB population present in plants from which the AY-WB DNA was isolated.

The variation among the AY-WB phytoplasmas suggests that they are prone to frequent mutations. This is consistent with other findings. OY-M has plasmids ranging from ~3 to ~ 7 kb in size (Fig 1B) (65), and the plasmids of beet leafhopper-transmitted virescence agent (BLTVA) phytoplasma range from ~2.5 to ~11 kb (50). There is high variability of the occurrence of ORFs in the plasmids of 30 beet leafhopper-transmitted virescence phytoplasma strains (50). There is also evidence of intramolecular recombination among phytoplasma plasmids (50, 65). We show that they can also recombine with the chromosome (Fig. 1C).

Phytoplasma metabolism. Except for a few exceptions described in Results, the AY-WB metabolic pathways are similar to those of OY-M that have been described elsewhere (68) and

will not be discussed in detail here though a few findings need more emphasis. The phytoplasma metabolism is in several ways different from those of SEM branch mollicutes. This was expected, because phytoplasmas have not been grown in cell-free culture medium, including mycoplasma culture media. Unlike SEM branch mollicutes, phytoplasmas do not have PTSs to import sugars and generate glucose-6-phosphate to feed the glycolysis pathway. Thus, phytoplasmas are clearly different from the insect-transmitted plant pathogenic *S. citri* and *S. kunkelii*, which have three PTSs for import of fructose, glucose and trehalose (André et al., 2003. Microbiology 149: 2687). In contrast, phytoplasmas possess ABC transporters for import of maltose. The maltose binding protein (MalE) (Table 4) may have affinity to maltose, trehalose, sucrose and palatinose (80). Affinity of MalE to trehalose is likely as trehalose is a major sugar in the insect hemolymph. The fate of these sugars after import is not clear, because enzymes required for conversion of these sugars to glucose-6-phosphate for glycolysis were not found in the phytoplasma genomes, and the sucrose phosphorylase gene, which is important for sucrose degradation is fragmented in the OY-M phytoplasma genome (68) and is completely absent from the AY-WB phytoplasma genome (Table 6). Generally, the genomes of AY-WB and OY-M phytoplasmas harbor significantly fewer carbohydrate transport and metabolism genes than their mycoplasma counterparts. Even in the 580-kb genome of *M. genitalium*, 26 carbohydrate transport and metabolism genes were identified (29). In contrast, only 19 genes are present in the 860-kb OY-M phytoplasma genome (68) and 16 genes in the 706-kb AY-WB phytoplasma genome.

Unlike SEM branch mollicutes, phytoplasmas have NAD-specific malic enzyme (EC: 1.1.1.38) and malate/citrate-sodium symporter genes. Thus, like symbiotic *Rhizobium* (69) but unlike sequenced SEM branch mollicutes, phytoplasmas may use malate as a carbon source. The

use of malate is advantageous, because it is readily available in the cytoplasm of host cells, and it can serve as the sole energy source for bacteria by conversion to oxaloacetate and pyruvate (23, 69). Further, metabolism of malate saves energy (23), which is important, because phytoplasmas lack ATP synthase and hence the capacity to generate energy in phytoplasmas seems limited to glycolysis (starting with glucose-6-phosphate).

Unlike SEM clade mollicutes, phytoplasmas appear to be capable of biosynthesis of their own membrane phospholipids. The genomes of AY-WB, OY-M (68) and Western X-disease phytoplasma (49) contain the *pssA* and *psd* genes (Table 6) encoding CDP-diacylglycerol-serine-O-phosphatidyltransferase (EC: 2.7.8.8) and phosphatidylserine decarboxylase (EC: 4.1.1.65), respectively. Both are part of the phosphatidylethanolamine (PE) pathway (58). Further, the AY-WB and OY-M genomes contain a candidate *pmt* gene for phospholipid N-methyltransferase (Table 6) that is involved in phosphatidylcholine (PC) synthesis in conjunction with PssA and Psd (58). This confirms that phytoplasmas are phylogenetically more related to acholeplasmas (4), which do not require exogenous phospholipids, whereas SEM branch mollicutes are sterol and fatty acid auxothrophs (72). AY-WB and OY-M also have all enzymes that link the glycolysis pathway to the glycerolipid pathway (68), and an ABC transporter gene *phnL* involved in lipoprotein release (Table 4).

Summary. Phytoplasmas have intriguing genomes that are small and contain many multicopy sequences mainly organized as PMUs. The AY-WB genome is ~154 kb smaller than the OY-M genome primarily as a result of fewer multicopy sequences. Thus, expansions or reductions of PMUs play a major role in phytoplasma genome evolution. At least one PMU, PMU1, has the characteristics of a replicative composite transposon. PMUs contain genes for specialized sigma factors and membrane proteins providing evidence that PMUs are important for phytoplasma

interactions with the environment. Since phytoplasmas lack *recA* and other standard homologous recombination functions, it is unlikely that phytoplasmas generate antigenic variation of membrane proteins through homologous recombination. Instead, we propose that expression regulation of PMU genes is one of the strategies phytoplasmas use to adapt to different environments. Expression of PMU genes might occur through a process that involves circularization and replicative transposition. In addition, genome rearrangements through expansions and deletions of PMUs might increase the chance of phytoplasma adaptation to diverse hosts, and can be a major evolutionary factor allowing phytoplasmas to occupy a broad plant host range or to adapt to different insect vectors. Few genes have similarities to known bacterial virulence factors. Like the related Gram-positive bacteria, phytoplasmas may secrete virulence-related proteins via the sec-dependent pathway. Hence, all the proteins with signal peptides are potential virulence factors, including the five solute binding proteins of the ABC transporters, and proteins derived from plasmids and PMUs. Finally, phytoplasmas have ABC transporters for the import of maltose (or trehalose, sucrose, palatinose), utilize malate, and can make phospholipids. In contrast, SEM branch mollicutes have PTSs for the import of fructose, glucose and trehalose, utilize lactate, and are phospholipid auxothrophs.

ACKNOWLEDGMENTS

This work was supported by the National Research Institute of the USDA Cooperative State Research, Education and Extension Service grant 2002-35600-12752, and the Ohio Agricultural Research and Development Center (OARDC) competitive grants program.

We thank members of the bioinformatics and genome analysis group at Integrated Genomics, including Svetlana Gerdes, Eugene Goltsman, Viktor Joukov, Vinayak Kapatral, Yakov Kogan,

Nikos Kyrpides, Andrei Osterman and Ross Overbeek. We also acknowledge Angela D. Strock, Melanie L. Lewis Ivey and Jhony Mera for excellent technical assistance.

REFERENCES

1. **Achaz, G., Coissac, E., Netter, P. and E. P. Rocha.** 2003. Associations between inverted repeats and the structural evolution of bacterial genomes. Genetics **164**:1279-1289.

2. **Akiyama, Y., Shirai, Y. K. and Ito K**. 1994. Involvement of FtsH in protein assembly into and through the membrane. II. Dominant mutations affecting FtsH functions. J. Biol. Chem. **269**:5225-5229.

3. **Allali, N., Afif, H., Couturier, M. and L. van Melderen**. 2002. The highly conserved TldD and TldE proteins of Escherichia coli are involved in microcin B17 processing and in CcdA degradation. J. Bacteriol. **184**:3224-3231.

4. **Allen, B.L. and M. Hook.** 2002. Isolation of a putative laminin binding protein from Streptococcus anginosus. Microb. Pathol. **33**:23-31.

5. **Alonso, J.C., Weise, F. and F. Rojo.** 1995. The Bacillus subtilis Histone-like Protein Hbsu Is Required for DNA Resolution and DNA Inversion Mediated by the β Recombinase of Plasmid pSM10935. J. Biol. Chem. **270**:2938-2945.

6. **Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and D.J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389-3402.

7. **Badger, J. H. and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16**:512-524.

8. **Bai, X., Zhang, J., Holford, I.R. and S. A. Hogenhout.** 2004. Comparative genomics identifies genes shared by distantly related insect-transmitted plant pathogenic mollicutes. FEMS Microbiol Lett. **235**:249-258.

9. **Barbara, D.J., Morton, A., Clark, M.F. and D. L. Davies**. 2002. Immunodominant membrane proteins from two phytoplasmas in the aster yellows clade (chlorante aster yellows and clover phyllody) are highly divergent in the major hydrophilic region. Microbiology **148**:157-167.

10. **Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khana, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. and S. R. Eddy.** 2004. The Pfam protein families database. Nucleic Acids Res. **32**:D138-D141.

11. **Beanland, L., Hoy, C.W., Miller, S.A. and L. R. Nault.** 2000. Influence of Aster Yellows Phytoplasma on the Fitness of Aster Leafhopper (Homoptera: Cicadellidae). Ann. Entomol. Soc. Am. **93**:271-276.

12. **Bendtsen, J.D., Nielsen, H., von Heijne, G. and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**: 783-795.

13. **Berducci, G., Mazzetti, A.P., Rotilio, G. and A. Battistoni.** 2004. Periplasmic competition for zinc uptake between the metallochaperone ZnuA and Cu,Zn superoxide dismutase. FEBS Lett. **569**:289-292.

14. **Boutareaud, A., Danet, J.L., Garnier, M. and C. Saillard C.** 2004. Disruption of a gene predicted to encode a solute binding protein of an ABC transporter reduces transmission of Spiroplasma citri by the leafhopper *Circulifer haematoceps*. Appl. Environ. Microbiol. **70**:3960-3967.

15. **Bové, J. M.** 1997. Spiroplasmas: Infectious agents of plants, arthropods and vertebrates. Wien. Klin. Wochenschr. **109**:604 – 612.

16. **Carr, K.M. and J. M. Kaguni.** 2002. *Escherichia coli* DnaA Protein Loads a Single DnaB Helicase at a DnaA Box Hairpin. J. Biol. Chem. **277**:39815-39822.

17. **Cheggour, A., Fanuel, L., Duez, C., Joris, B., Bouillenne, F., Devreese, B., Van Driessche, G., Van Beeumen, J., Frere, J.M. and C. Goffin.** 2000. The dppA gene of *Bacillus subtilis* encodes a new D-aminopeptidase. Mol Microbiol. **38**:504-513.

18. **Czosnek, H., Ghanim, M., Morin, S., Rubinstein, G., Fridman, V. and M. Zeidan**. 2001. Whiteflies: vectors, and victims (?), of geminiviruses. Adv Virus Res. **57**:291-322.

19. **Davidson, A.L. and J. Chen.** 2004. ATP-binding cassette transporters in bacteria. Annu. Rev. Biochem. **73**:241-268.

20. **Davis, R.E., Jomantiene, R., Zhao, Y., and E. L. Dally.** 2003. Folate biosynthesis pseudogenes, PsifolP and PsifolK, and an O-sialoglycoprotein endopeptidase gene homolog in the phytoplasma genome. DNA Cell Biol. **22**:697-706.

21. **Delcher, A. L., Harmon, D., Kasif, S., White, O. and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27**:4636-4641.

22. **Deuerling, E., Mogk, A., Richter, C., Purucker, M. and W. Schumann.** 1997. The ftsH gene of *Bacillus subtilis* is involved in major cellular processes such as sporulation, stress adaptation and secretion. Mol Microbiol. **23**:921-933.

23. **Dimroth, P. and B. Schink.** 1998. Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. Arch Microbiol. **170**:69-77.

24. **Eisen, J.A., Heidelberg, J.F., White, O. and S. L. Salzberg.** 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. **1**: research0011.1–0011.9

25. **Ewing, B., Hillier, L., Wendl, M.C. and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**:175-185.

26. **Ewing, B. and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186-194.

27. **Fernandez, S. Rojo, F. and J. C. Alonso.** 1997. The *Bacillus subtilis* chromatin-associated protein Hbsu is involved in DNA repair and recombination. Mol. Microbiol. **23**:1169-1179.

28. **Frank, A.C. and J. R. Lobry.** 2000. Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. Bioinformatics **16**: 560-561.

29. **Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, R. D., Weidman, J. F., Small, K. V., Sandusky, M. Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith H. O., Hutchison, C. A. 3rd, and J. C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science*. **270**: 397-403.

30. **Gasparich, G.E.** 2002. Spiroplasmas: evolution, adaptation and diversity. Front Biosci. **7**:d619-40.

31. **Guy, L. and C. A. Roten.** 2004. Genometric analyses of the organization of circular chromosomes: a universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. Gene **340**:45-52.

32. **Hall, M.C., Jordan, J.R. and S. W. Matson**. 1998. Evidence for a physical interaction between the *Escherichia coli* methyl-directed mismatch repair proteins MutL and UvrD. EMBO J. **17**:1535-1541.

33. **Hanboonsong, Y., Choosai, C., Panyim, S. and S. Damak.** 2002. Transovarial transmission of sugarcane white leaf phytoplasma in the insect vector *Matsumuratettix hiroglyphicus* (Matsumura). Insect Mol Biol. **11**:97-103.

34. **Higgins, C.F**. 2001. ABC transporters: physiology, structure and mechanism – an overview. Res. Microbiol. **152**:205-210.

35. **Hosie, A.H.F. and P. S. Poole.** 2001. Bacterial ABC transporters of amino acids. Res. Microbiol. **152**:259-270.

36. **Hruska, A. J., and M. Gomez Peralta**. 1997. Maize Response to Corn Leafhopper (Homoptera: Cicadellidae) Infestation and Achaparramiento Disease. J. Econ. Entomol. **90**: 604-610

37. **Jaffe, J, D., Stange-Thomann, N., Smith, C., DeCaprio, D., Fisher, S., Butler, J., Calvo, S., Elkins, T., FitzGerald, M.G., Hafez, N., Kodira, C. D., Major, J., Wang, S., Wilkinson, J., Nicol, R., Nusbaum, C., Birren, B., Berg, H. C. and G. M. Church.** 2004. The Complete Genome and Proteome of *Mycoplasma mobile*. *Genome Research* **14**: 1447-1461.

38. **Kakizawa, S., Oshima, K., Kuboyama, T., Nishigawa, H., Jung, H., Sawayanagi, T., Tsuchizaki, T., Miyata, S., Ugaki, M. and S. Namba.** 2001. Cloning and expression analysis of Phytoplasma protein translocation genes. Mol Plant Microbe Interact. **14**:1043-1050.

39. **Kakizawa, S., Oshima, K., Nishigawa, H., Jung, H.Y., Wei, W., Suzuki, S., Tanaka, M., Miyata, S., Ugaki, M. and S. Namba.** 2004. Secretion of immunodominant membrane protein from onion yellows phytoplasma through the Sec protein-translocation system in *Escherichia coli*. Microbiology **150**:135-142.

40. **Khan, S.A.** 1997. Rolling-Circle Replication of Bacterial Plasmids. Microbiol. Mol. Biol. Rev. **61**:442-455.

41. **Kieboom, L., Bruinenberg, R., Keizer-Gunnink, I. and J. A. de Bont.** 2001. Transposon mutations in the flagella biosynthetic pathway of the solvent-tolerant *Pseudomonas putida* S12 result in a decreased expression of solvent efflux genes. FEMS Microbiol Lett. **198**: 117-122.

42. **King, K.W., Woodard, A. and K. Dybvig.** 1994. Cloning and characterization of the *recA* genes from *Mycoplasma pulmonis* and *M. mycoides* subsp. *mycoides*. Gene **139**:111-115.

43. **Kiss, J. and F. Olasz.** 1999. Formation and transposition of the covalently closed IS30 circle: the relation between tandem dimmers and monomeric circles. Mol. Microbiol. **34**:37-52.

44. **Koonin, E.V. and T. V. Ilyina.** 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol. **73**:2763-2766.

45. **Kormanec, J. Seccikova, B., Halgasova, N., Knirschova, R. and B. Rezuchova.** 2000. Identification and transcriptional characterization of the gene encoding the stress-response sigma factor sigma(H) in *Streptomyces coelicolor* A3(2). FEMS Microbiol. Lett. **189**:31-38.

46. **Lee, I.M., Davis, R.E. and D. E. Gundersen-Rindal**. 2000. Phytoplasma: phytopathogenic mollicutes. Annu Rev Microbiol. **54**:221-55.

47. **Lee, I-M., Gundersen-Rindal, D. E., Davis, R.E., Bottner, K.D., Marcone, C. and E. Seemuller.** 2004. '*Candidatus* Phytoplasma asteris', a novel phytoplasma taxon associated with aster yellows and related diseases. Int. J. Syst. Evol. Microbiol. **54**: 1037-1048.

48. **Lee, I.M., Zhao, Y. and K. D. Bottner.** 2005. Novel insertion sequence-like elements in phytoplasma strains of the aster yellows group are putative new members of the IS3 family. FEMS Microbiol Lett. **242**:353-360.

49. **Liefting, L.W. and B. C. Kirkpatrick.** 2003. Cosmid cloning and sample sequencing of the genome of the uncultivable mollicute, Western X-disease phytoplasma, using DNA purified by pulsed-field gel electrophoresis. FEMS Microbiology Letters **221**:203-211.

50. **Liefting, L. W., Shaw, M. E. and B. C. Kirkpatrick**. 2004. Sequence analysis of two plasmids from the beet leafhopper-transmitted virescence agent. Microbiology **150**:1809-1817.

51. **Lohman, T.M. and M. E. Ferrari.** 1994. *Escherichia coli* single-stranded DNA-binding protein: Multiple DNA-Binding Modes and Cooperativities. Annu. Rev. Biochem. **63**:526-570.

52. **Lysenko, E., Ogura, T. and S. M. Cutting.** 1997. Characterization of the *ftsH* gene of *Bacillus subtilis*. Microbiology **143**:971-978.

53. **Mahillon, J. and M. Chandler.** 1998. Insertion Sequences. Microbiol. Mol. Biol. Rev. **62**:725-774.

54. **Marais, A., Bove, J.M. and J. Renaudin**. 1996. Characterization of the *recA* gene regions of *Spiroplasma citri* and *Spiroplasma melliferum*. J. Bacteriol. **178**:7003-7009.

55. **Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., Liebert, C. A.,**

Liu, C., Madej, T., Marchler, G. H., Mazumder, R., Nikolskaya, A. N., Panchenko, A. R., Rao, B. S., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Vasudevan, S., Wang, Y., Yanashita, R. A., Yin, J. J. and S. H. Bryant. 2003. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res. **31**:383-387.

56. **Marcone, C., Neimark, H., Ragozzino, A., Lauer, U. and E. Seemüller.** 1999. Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups. Phytopathology **89**:805-810.

57. **Marcone, C., I. M. Lee, R. E. Davis, A. Ragozzino, and E. Seemuller.** 2000. Classification of aster yellows-group phytoplasmas based on combined analyses of rRNA and tuf gene sequences. Int J Syst Evol Microbiol **50**:1703-1713.

58. **Martinez-Morales, F., Schobert, M., Lopez-Lara, I.M. and O. Geiger.** 2003. Pathways for phophatidylcholine biosynthesis in bacteria. Microbiology **149**:3461-3471.

59. **Mendonca, V.M. and S. W. Matson**. 1995. Genetic Analysis of Δ*helD* and Δ*uvrD* Mutations in Combination with Other Genes in the RecF Recombination Pathway in Escherichia coli: Suppression of *ruvB* Mutation by a *uvrD* Deletion. Genetics **141**:443-452.

60. **Miyata, S., Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.Y., Kuboyama, T., Ugaki, M. and S. Namba.** 2003. Two different thymidylate kinase gene homologues, including one that has catalytic activity, are encoded in the onion yellows phytoplasma genome. Microbiology **149**:2243-2250.

61. **Montano, H.G., Davis, R.E., Dally, E.L., Hogenhout, S., Pimentel, J.P. and P. S. T. Brioso.** 2001. 'Candidatus Phytoplasma brasiliense', a new phytoplasma taxon associated with hibiscus witches' broom disease. Int. J. Syst. Bacteriol. **51**:1109-1118.

62. **Murayama, N., Shimizu, H., Takiguchi, S., Baba, Y., Amino, H., Horiuchi, T., Sekimizu, K. and T. Miki.** 1996. Evidence for involvement of *Escherichia coli* genes *pmbA, csrA* and a previously unrecognized gene *tldD*, in the control of DNA gyrase by *letD* (*ccdB*) of sex factor F. J Mol Biol. **256**:483-502.

63. **Nishigawa, H., Miyata, S., Oshima, K., Sawayanagi, T., Komoto, A., Kuboyama, T., Matsuda, I., Tsuchizaki, T. and S. Namba.** 2001. *In planta* expression of a protein encoded by the extrachromosomal DNA of a phytoplasma and related to geminivirus replication proteins. Microbiology **147**:507-513.

64. **Nishigawa, H., Oshima, K., Kakizawa, S., Jung, H.Y., Kuboyama, T., Miyata, S., Ugaki, M. and S. Namba.** 2002. 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 **298**:195-201.

65. **Nishigawa, H., Oshima, K., Kakizawa, S., Jung, H.Y., Kuboyama, T., Miyata, S., Ugaki, M. and S. Namba.** 2002. Evidence of intermolecular recombination between extrachromosomal DNAs in phytoplasma: a trigger for the biological diversity of phytoplasma? Microbiology **148**:1389-1396.

66. **Olasz, F., Farkas, T., Kiss, J., Arini, A. and W. Arber.** 1997. Terminal Inverted Repeats of Insertion Sequence IS*30* Serve as Targets for Transposition. J. Bacteriol. **179**:7551-7558.

67. **Oshima, K., Kakizawa, S., Nishigawa, H., Kuboyama, T., Miyata, S., Ugaki, M. and S. Namba.** 2001. A plasmid of phytoplasma encodes a unique replication protein having both plasmid- and virus-like domains: clue to viral ancestry or result of virus/plasmid recombination? Virology **285**:270-277.

68. **Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.Y., Wei, W., Suzuki, S., Arashida, R., Nakata, D., Miyata, S., Ugaki, M. and S. Namba.** 2004. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. Nat Genet. **36**:27-29.

69. **Poole, P. and D. Allaway.** 2000. Carbon and nitrogen metabolism in Rhizobium. Adv Microb Physiol. **43**:117-163.

70. **Prévost, C. and M. Takahashi.** 2004. Geometry of the DNA strands within the RecA nucleofolament: role in homologous recombination. Q. Rev. Biophys. **36**:429-453.

71. **Ray, K.C., Tu, Z-C., Grogono-Thomas, R., Newell, D.G., Thompson, S.A. and M. J. Blaser.** 2000. *Campylobacter fetus* Inversion Occurs in the Absence of RecA Function. Inf. Immun. **68**:5663-5667.

72. **Razin, S., Yogev, D. and Y. Naot.** 1998. Molecular Biology and Pathogenicity of Mycoplasmas. Microbiol. Mol. Biol. Rev. **62**:1094-1156.

73. **Rekab, D., Carraro, L., Schneider, B., Seemuller, E., Chen, J., Chang, C.J., Locci, R. and G. Firrao.** 1999. Geminivirus-related extrachromosomal DNAs of the X-clade phytoplasmas share high sequence similarity. Microbiology **145**:1453-1459.

74. **Rocha, E.P. and A. Blanchard.** 2002. Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. Nucleic Acids Res. **30**:2031-2042.

75. **Rodriguez-Sainz, M.C., Hernandez-Chico, C. and F. Moreno**. 1990. Molecular characterization of pmbA, an *Escherichia coli* chromosomal gene required for the production of the antibiotic peptide MccB17. Mol Microbiol. **4**:1921-1932.

76. **Rosch, J. and M. Caparon.** 2004. A Microdomain for Protein Secretion in Gram-positive Bacteria. Science **304**:1513-1515.

77. **Schneider, B., Torres, E., Martin, M.P., Schroder, M., Behnke, H.D. and E. Seemuller.** 2005. '*Candidatus* Phytoplasma pini', a novel taxon from *Pinus silvestris* and *Pinus halepensis*. *Int* J Syst Evol Microbiol. **55**:303-307.

78. **Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and D. Schomburg**. 2004. BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res. **32**:D431-D433.

79. **Segal, G., Feldman, M. and T. Zusman.** 2005. The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and *Coxiella burnetii*. FEMS Microbiol Rev. **29**:65-81.

80. **Silva, Z., Sampaio, M.M., Henne, A., Bohm, A., Gutzat, R., Boos, W., da Costa, M.S. and H. Santos.** 2005. The high-affinity maltose/trehalose ABC transporter in the extremely thermophilic bacterium *Thermus thermophilus* HB27 also recognizes sucrose and palatinose. J Bacteriol. **187**:1210-8.

81. **Szeverényi, I., Nagy, Z., Farkas, T., Olasz, F. and J. Kiss**. 2003. Detection and analysis of transpositionally active head-to-tail dimmers in three additional *Escherichia coli* IS elements. Microbiology **149**:1297-1310.

82. **Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A. and S. G. Andersson.** 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296**:2376-2379.

83. **Tettelin, H., Radune,D., Kasif,S., Khouri,H.and S. L. Salzberg**. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. Genomics **62**:500-507

84. **The IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group**. 2004. '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int. J. Syst. Evol. Microbiol. **54**:1243–1255.

85. **Tougu K. and K. J. Marians**. 1996. The Interaction between Helicase and Primase Sets the Replication Fork Clock. J. Biol. Chem. **35**:21398-21405.

86. **Weisburg, W. G., Tully, J. G., Rose, D. L., Petzel, J. P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T. G., Van Etten, J., Maniloff, J. and C. R. Woese.** 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. **171**:6455-6467.

87. **Westberg, J., Persson, A., Holmberg, A., Goesmann, A., Lundeberg, J., Johanssen, K-E., Petterson, B. and M. Uhlén.** 2004. The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* type strain PG1^{T}, the causative agent of contagious bovine pleuropneumonia (CPBB). Genome Research **14**:221-227.

88. **Woese, C.R**. 1987. Bacterial evolution. Microbial Rev. **51**:221-227.

89. **Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W., Esser, C., Ahmadinejad, N., Wiegand, C., Madupu, R., Beanan, M. J., Brinkac, L. M., Daugherty, S. C., Durkin, A. S., Kolonay, J. F., Nelson, W. C., Mohamoud, Y., Lee, P., Berry, K., Young, M. B., Utterback, T., Weidman, J., Nierman, W. C., Paulsen, I. T., Nelson, K. E., Tettelin, H., O'Neill, S. L. and J. A. Eisen.** 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. **E69**.

90. **Yu, F., Inouye, S. and M. Inouye.** 1986. Lipoprotein-28, a Cytoplasma Membrane Lipoproteins from *Escherichia coli*. J. Biol. Chem. **261**:2284-2288.

91. **Zhang, J., Hogenhout, S. A., Nault, L.R., Hoy, C.W. and S. A. Miller.** 2004. Molecular and symptom analyses of phytoplasma strains from lettuce reveal a diverse population. Phytopathology **94**:842-849

1 TABLE 1. General features of the chromosomes of AY-WB and OY-M.

	AY-WB	OY-M[a]
Length (bp)	706,569	860,631
G + C content (percent)	27	28
Protein-coding region (percent)	72	73
Protein-coding genes with assigned function	450	446
Conserved hypothetical	149[b]	51
Hypothetical	72	257
Total	671	754
Average length of protein-coding genes (bp)	779	785
tRNA	31[c]	32[c]
rRNA operons	2	2

2 [a] Numbers taken from Oshima et al., 2004 (68).

3 [b] Includes proteins with similarity (pblast < 10^{-5}) to OY-M proteins.

4 [c] tRNA corresponding to all amino acids are represented.

TABLE 2. General features of the plasmids of AY-WB and OY-M.

	AYWB pI	AYWB pII	AYWB pIII	AYWB pIV	EcOYM[a]	pOYM[a]
Length (bp)	3,972	4,009	5,104	4,316	5,025	3,932
G + C content (percent)	25.6	23.9	21.8	25.5	25	24
Protein-coding region (percent)	75	71	65	76	71	75
Protein-coding genes with assigned function	2	2	2	2	2	2
Conserved hypothetical	3[b]	2[b]	6[b]	3[b]	-	-
Hypothetical	-	-	1	1	4	3
Total	5	4	7	6	6	5
Average length of protein-coding genes (bp)	594	569	472	546	597	588

[a] Numbers taken from Oshima et al., 2004 (68).

[b] Includes proteins with similarity (pblast < 10^{-5}) to OY-M proteins.

TABLE 3. Features of the four potential mobile units (PMUs) of AY-WB.

ORF[a]	ORF ID				Annotation
	PMU1	**PMU2**	**PMU3**	**PMU4**	
1	*tra5* (210)[c]	-	-	-	Truncated transposase, group IS*150*, family IS*3*
2	*sigF* (624)	*sigF* (603)	-	-	Specialized sigma factor
3	*ssb* (312)	*ssb* (333)	*ssb* (312)	-	Single-stranded DNA binding protein
4	*himA* (330)	*himA* (288)	*himA* (366)	-	DNA-binding factor HU
5	*AYWB_191* (438)	-	*AYWB_273* (441)	-	Cons hyp protein
6[b, f]	*AYWB_190* (279)	-	*AYWB_274* (294)	-	Hyp protein
7[b]	*AYWB_189* (792)	-	-	-	Cons hyp protein
8[b]	*AYWB_188* (858)	-	-	-	Cons hyp protein
9[b]	*hflB* (2,106)	*hflB* (2,304)	*hflB* (2,145)	-	Zn-dependent protease
10[b,f]	*AYWB_186* (270)	-	-	-	Cons hyp protein
11[b]	*AYWB_185* (855)	-	-	-	Cons hyp protein
12[e]	*AYWB_184* (2,253)	*AYWB_226* (618)	[d]*AYWB_277* (1,155); *AYWB_278* (1,110)	-	Cons hyp. protein
13[b]	*AYWB_183* (987)	*AYWB_225* (690)	*AYWB_279* (804)	-	Cons hyp protein
14	*AYWB_182* (636)	*AYWB_224* (372)	*AYWB_281* (366)	-	Cons hyp. protein
15	*tmk-a* (630)	*tmk-a* (630)	*tmk-a* (627)	-	Thymidylate kinase
16	*AYWB_180* (609)	*AYWB_221* (603)	*AYWB_283* (609)	*AYWB_618* (744)	Cons hyp. protein
17	*dnaB* (1,494)	*dnaB* (1,494)	*dnaB* (1,500)	*dnaB* (1,413)	DNA helicase
18	*dnaG* (1,323)	*dnaG* (1,323)	*dnaG* (1,323)	*dnaG* (1,107)	DNA primase
19[b]	*AYWB_177* (855)	*AYWB_218* (162)	-	*AYWB_615* (834)	Cons hyp protein
20	*AYWB_176* (624)	[d]*AYWB_217* (312); *AYWB_216* (360)	*AYWB_286* (750)	*AYWB_614* (564)	Cons hyp. protein
21	*tra5* (963)	*tra5* (939)	*tra5* (963)	*tra5* (396, 519)[e]	Transposase, group IS*150*, family IS*3*
22[f]	-	*AYWB_231* (171)	-	-	Hyp. protein
23[b]	-	*AYWB_228* (873)	*AYWB_276* (600)	-	Cons hyp protein
24	-	*AYWB_227 (411)*	-	-	Cons hyp protein
25[b]	-	*AYWB_223* (627)	-	-	Cons hyp protein
26[b]	-	-	*AYWB_280* (261)	-	Cons hyp protein
27	-	-	*mgs1* (1,242)	-	ATPase, AAA family
28	-	-	*tra5* (963)	-	Transposase, group IS*150*, family IS*3*

[a] ORF numbers corresponding to numbers of Fig. 3.

[b] Deduced proteins predicted to target the membrane (secreted or membrane proteins).

[c] ORF IDs with lengths in nucleotides between brackets are indicated for all PMU ORFs.

[d] Genes contain mutations separating them in two truncated ORFs (Fig 3).

[e] Contains separate A and B ORFs that may produce a full-length transposase upon a single frameshift event (53).

[f] Sequences unique to AY-WB.

[e] Sequences conserved among most mollicutes. All other conserved hypothetical proteins are conserved solely between AY-WB and OY-M.

Abbreviations: Cons, conserved; hyp, hypothetical.

TABLE 4. Summary of ABC transporter genes in AY-WB and OY-M phytoplasma genomes

Substrate	AY-WB ATP-binding protein	AY-WB Membrane protein	AY-WB Solute-binding protein	OY-M ATP-binding protein	OY-M Membrane protein	OY-M Solute-binding protein
Amino acid uptake						
Amino acid	*glnQ* (AYWB_634)	*AYWB635* (AYWB_635)		*glnQ* (39938565),	*artM* (39938563), *artM* (39938564)	
D-methionine	*metN* (AYWB_589)	*AYWB587* (AYWB_587)	*nlpA* (AYWB_588)	*abc* (39938618)	*PAM134* (39938620)	*nlpA* (39938619)
Amino acid (arginine)	AYWB_314 (fragment)	*glnP* (AYWB_315)			*artM* (39938942), *artI* (39938943), *artM* (39938950)	
Amino acid (glutamine)	*artP* (AYWB_264)	*artQ* (AYWB_265), *artM* (AYWB_262)	*artI* (AYWB_263)	*glnQ* (39938974)	*artM* (39938973), *artI* (39938975), *artM* (39938976)	
Amino acid		*artM* (AYWB_125)			*artM* (39939074)	
Amino acid					*artM* (39938980), *artM* (39938981)	
Amino acid					*artM* (39939125), *mdoB* (39939127)	
Dipeptide/oligopeptide uptake						
Dipeptide or oligopeptide	*dppF* (AYWB_527) *dppD* (AYWB_528)	*dppB* (AYWB_530), *dppC* (AYWB_531)	*dppA* (AYWB_529)	*dppD* (39938678)	*dppC* (39938675), *dppB* (39938676)	*oppA* (39938677)
Oligopeptide				*dppD* (39938511), *oppF* (39938512)	*dppB* (39938508), *PAM023* (39938509)	*PAM024* (39938510)
Sugar uptake						
Maltose, trehalose, sucrose or palatinose	*malK* (AYWB_670)	*malG* (AYWB_668), *malF* (AYWB_669)	*malE* (AYWB_667)	*malK* (39939238)	*ugpE* (39939236), *ugpA* (39939237)	*ugpB* (39939235)
Inorganic ion uptake						
Cobalt	*cbiO* (AYWB_014)	*cbiQ* (AYWB_015)		*cbiO* (39938506)	*PAM19* (39938505)	
Cobalt	*cbiO* (AYWB_540) *cbiO* (AYWB_541)	*cbiQ* (AYWB_539)		*cbiO* (39938665)	*cibQ* (39938666)	
Mn/Zn	*mntA* (AYWB_623)	*mntB* (AYWB_622), *mntB* (AYWB_621)	*znuA* (AYWB_624)	*znuC* (39938579)	*znuB* (39938580)	*znuA* (39938578)
Multidrug resistence						
Multidrug	*evbG/mdlB* (AYWB_028)			*mdlB* (39938545)		
Multidrug	*evbH* (AYWB_029)					
Spermidine/putrescine uptake						
Spermidine or putrescine	*potA* (AYWB_095)	*potB* (AYWB_094), *potC* (AYWB_093)	*potD* (AYWB_092)	*potA* (39939145)	*potB* (39939146), *potC* (39939147)	*potD* (39939148)
Uncharacterized						
Possible lipoprotein	*phnL* (AYWB_619)			*phnL* (39938582)		*nlpA* (39938583)
Unknown	*phnL* (AYWB_135)			*phnL* (39939085)		

TABLE 5. Predicted P-type ATPases of AY-WB and OY-M

AYWB		OY-M	
Gene (Length, CDs)	**Possible substrate**	**Gene (Length, Acc. no.)**	**Possible substrate**
mgtA (920 aa, AYWB_018)	Cation	*mgtA* (920 aa, 39938516)	Sodium/potassium
mgtA (817 aa, AYWB_469)	Cation	*mgtA* (918 aa, 39938672)	Calcium
mgtA (952 aa, AYWB_533)	Cation	*mgtA* (1056 aa, 39938738)	Cation
mgtB (892 aa, AYWB_242)	Magnesium	*mgtA* (892 aa, 39939071)	Magnesium
zntA (666 aa, AYWB_650)	Lead, cadmium, zinc, mercury	*zntA* (666 aa, 39939219)	Cadmium

TABLE 6. Proteins with functional annotations unique to AY-WB and OY-M within the class Mollicutes

AYWB ORF ID	Gene ID	Annotation	OY-M		Other organisms	
			GenBank Acc.[a]	E-value[b]	GenBank Acc.[a]	E-value[b]
Transcription						
AYWB_654	*rpoZ*	EC 2.7.7.6	39939222	$9e^{-19}$	58337597	$2e^{-07}$
Translation						
AYWB_504	*rpmD*	LSU ribosomal protein L30P	39938704	$5e^{-28}$	50590420	$3e^{-09}$
Membrane transport						
AYWB_052	*citS*	Malate-sodium symporter	39939206	e^{-166}	42528200	$5e^{-36}$
AYWB_435	*citS*	Malate-sodium symporter	39938772	e^{-174}	15672883	$2e^{-30}$
AYWB_125	*artM*	ABC-type permease protein ArtM	39939074	0	48866203	$2e^{-35}$
AYWB_263	*artI*	ABC type solute-binding protein ArtI	39938975	$6e^{-60}$	58336459	$1e^{-21}$
AYWB_265	*artQ*	ABC-type permease protein ArtQ	39938973	$9e^{-79}$	24376619	$2e^{-17}$
AYWB_315	*glnP*	ABC-type permease protein GlnP	39938942	0	15022937	$2e^{-30}$
AYWB_587		ABC-type Met ATP-binding protein	39938620	e^{-101}	29377647	$1e^{-13}$
AYWB_588	*nlpA*	ABC-type Met binding protein	39938619	e^{-136}	25010851	$3e^{-05}$
AYWB_621	*mntB*	ABC-type membrane protein	39938580	e^{-177}	42526732	$2e^{-53}$
AYWB_622	*mntB*	ABC-type membrane protein	39938580	e^{-162}	53685687	$1e^{-47}$
AYWB_624	*znuA*	ABC type Mn/Zn-binding protein	39938578	e^{-168}	1335912	$1e^{-41}$
AYWB_667	*malE*[a]	ABC type maltose-binding protein	39939235	0	52858068	$2e^{-18}$
AYWB_439	*norM*	Na+ driven multidrug efflux pump	39938768	0	n/a	n/a
AYWB_441	*norM*	Na+ driven multidrug efflux pump	39938766	0	n/a	n/a
AYWB_467	*secE*[a]	SecE	40786355	$1e^{-37}$	n/a	n/a
Metabolic enzymes						
AYWB_051	*sfcA*	EC 1.1.1.38	39939207	0	28202548	e^{-129}
AYWB_120	*pssA*	EC 2.7.8.8	39939099	$8e^{-92}$	15023686	$2e^{-16}$
AYWB_121	*psd*	EC 4.1.1.65	39939098	e^{-178}	15023687	$9e^{-50}$
AYWB_326	*sodA*	EC 1.15.1.1	39938928	e^{-107}	15672390	$1e^{-60}$
AYWB_415	*pmt*	EC 2.1.1.-	39938792	0	45682627	$4e^{-06}$
AYWB_470	*pnp*[a]	EC 2.7.7.8	39938737	0	48824146	e^{-178}
AYWB_532		EC 3.4.17.19	39938673	0	52698549	e^{-146}
AYWB_598	*qns*[a]	EC 6.3.5.1	39938607	0	16804107	e^{-158}
AYWB_607	*pcnB*	EC 2.7.7.19	39938586	$2e^{-14}$	n/a	n/a
Other						
AYWB_017	*ibpA*	Hsp20	39938514	$9e^{-64}$	4884483	$7e^{-14}$
AYWB_302	*mutT*	Phosphohydrolase	39938962	$8e^{-92}$	15673603	$2e^{-24}$
AYWB_331	*tldD*	TldD	39938933	0	15024804	e^{-107}
AWYB_332	*pmbA*	PmbA	39938933	0	18143998	$2e^{-54}$
AYWB_561	*hlyC*[a]	HemolysinIII	39938644	e^{-110}	18145579	$5e^{-26}$
AYWB_599		Immunodominant protein precursor (remove EC 6.3.5.1)	39938608	$2e^{-10}$	n/a	n/a
AYWB_630		Rhodanese-related sulfurtransferase	39938571	e^{-180}	23098027	e^{-109}
AYWB_646	*pduL*	PduL	39939215	e^{-102}	49235943	$5e^{-44}$

[a]Genes with identical annotations but no sequence similarities in SEM branch mollicute.

[b]The E-values were obtained by searching against GenBank non-redundant database with 2,506,223 sequences consisting of 849,940,114 letters on a local Linux workstation. Results from the GenBank search were verified using the mollicute database MolliGen (http://cbi.labri.fr/outils/molligen/) (Barre et al., 2004; Nucleic Acids Res. 1;32(Database issue):D307-10).

Figures and Figure legends:

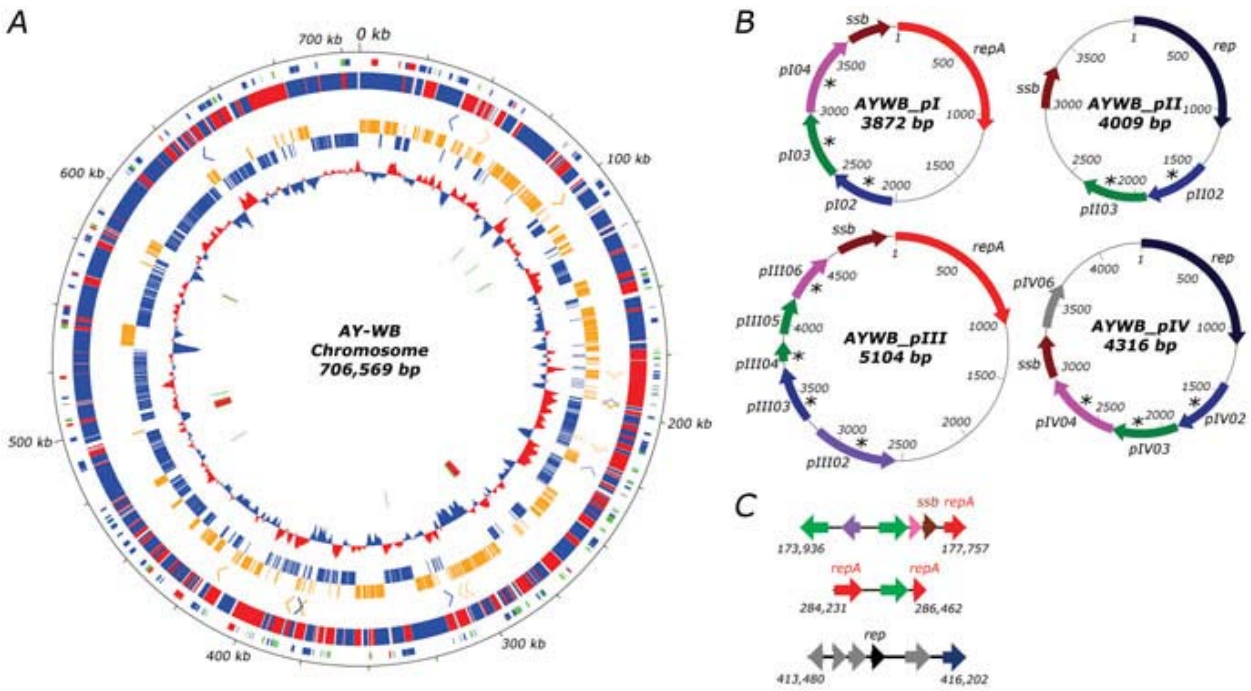


FIG. 1. (A) Genome maps of the 706,569 bp circular chromosome of *Candidatus* Phytoplasma asteris strain AY-WB. Rings present from the inside to outside: Ring 1, *rrn* operons in red and tRNA in green; Ring 2, GC skew over a 2-kb window and 200 bp steps with red denoting G > C and blue C > G; Ring 3, predicted ORFs in sense orientation in yellow and antisense orientation in blue; Ring 4, location of *tra5* ISs presented as angular brackets with yellow indicating sense orientation and blue antisense orientation; Ring 5, ORFs present in all sequenced mollicutes in blue and unique to phytoplasmas within the class Mollicutes in red; Ring 6, ORFs of predicted secreted proteins in green, secreted membrane proteins in red, and membrane proteins in blue; Ring 7, bp indicator with the first nucleotide of *dnaA* as nucleotide 1. The *oriC* is most likely located immediately upstream of *dnaA* as predicted by the Oriloc software (28), and the opposite direction of ORFs surrounding the putative *oriC*. (B) The four plasmids of AY-WB. ORFs are presented as block arrows with names of deduced protein sequences on the outside of the rings. Numbers on the inside of the rings indicate location in bp

with the first nucleotide of the *repA* and *rep* genes as nucleotide 1. ORFs indicated with * are predicted to encode membrane-targeted proteins. (C) Three chromosomal segments containing ORFs with similarity to plasmid ORFs. The chromosome is presented as a black line. The numbers below the black lines indicate the positions of the first and last nucleotide of the sequence on the AY-WB chromosome in bp. ORFs are represented as block arrows. Arrows of paralogous genes on plasmids and chromosome have the same color with exception of the grey-colored arrows, which represent unique genes. The names of the ORFs with predicted functions are indicated above the arrows. RepA, plasmid replication associated protein with significant similarity to RepA of geminiviruses. Rep, phytoplasma-specific plasmid replication protein. ssb, single-stranded DNA-binding protein.

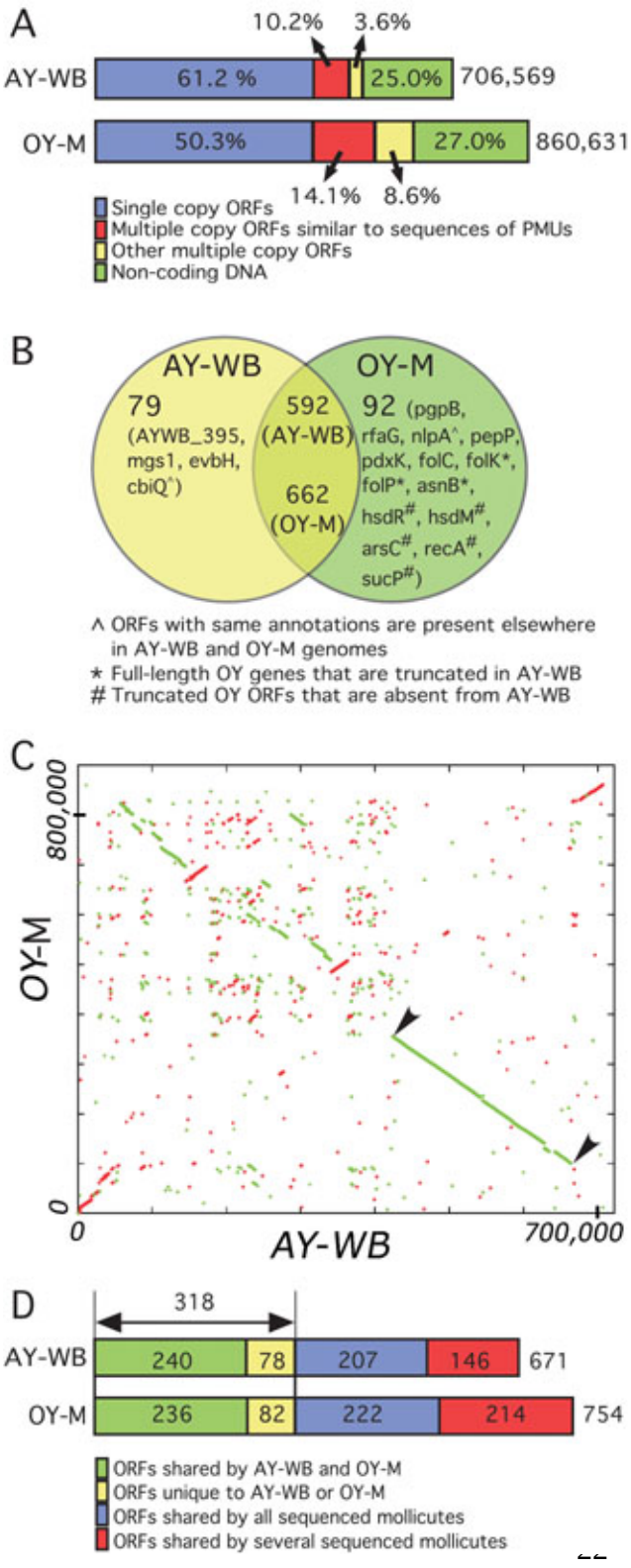


FIG. 2. Comparative genome analysis of the AY-WB genome with the genomes of OY-M and other mollicutes. (A) The AY-WB and OY-M genomes are repeat-rich. PMUs, putative mobile units (Fig. 3). (B) Venn Diagram showing the number of shared and unique genes of AY-WB and OY-M. (C) Dotplot comparison of AY-WB and OY-M chromosomes. The numbers on the x- and y-axis indicate the nucleotides in bp. AY-WB and OY-M genome segments in the same orientation are represented as red lines, and those in the reverse orientation as green lines. The arrowheads indicate *lplA* and *glnQ* that flank ~250 kb of sequences mostly conserved among mollicutes. (D) The number of ORFs unique to phytoplasmas or shared with sequenced SEM clade mollicutes based on blastp analysis of AY-WB and OY-M protein sequences against a database composed of deduced protein

23 sequences of all fully sequenced mollicute genomes (E-value $<10^{-5}$). Accession numbers:

Mesoplasma florum L1 (AE017263); *M. gallisepticum* R. (AE015450), *M. genitalium* G-37 (L43967), *M. hyopneumoniae* 232 (AE017332); *M. mobile* 163K (AE017308), *M. mycoides* subsp. *mycoides* SC str. PG1 (BX293980), *M. penetrans* HF-2 (BA000026), *M. pneumoniae* M129 (U00089), *M. pulmonis* UAB CTIP (AL445566), OY-M phytoplasma (AP006628) and *U. urealyticum* serovar 3 str. ATCC (AF222894).

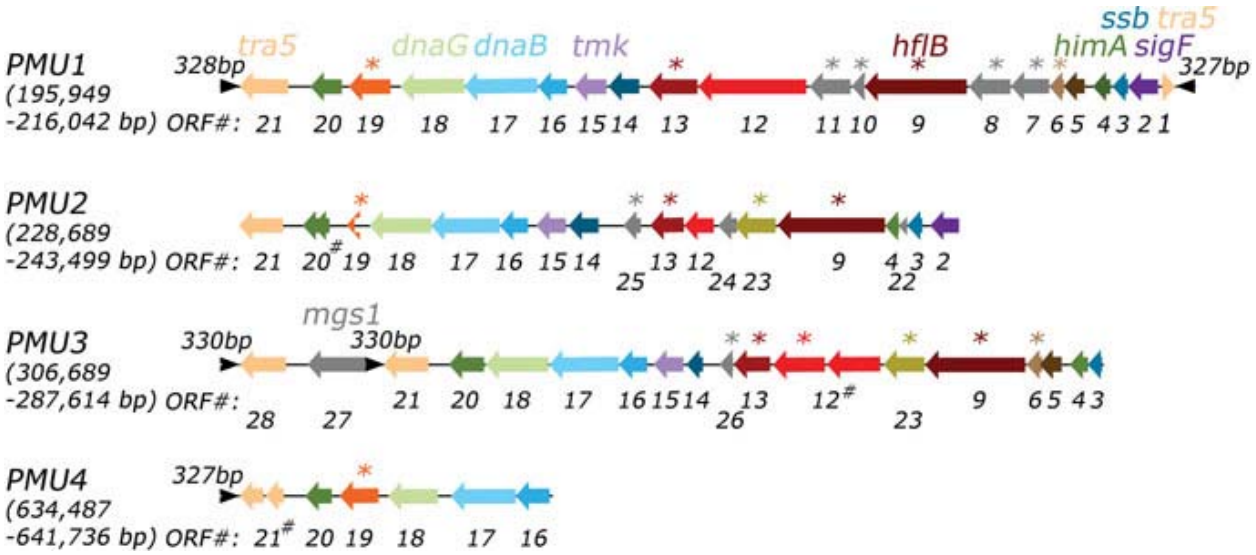


FIG. 3. Potential mobile units (PMUs) of the AY-WB chromosome. The chromosome is presented as a black line. The numbers between brackets at the left indicate the positions of the first and last nucleotide of the PMU on the AY-WB chromosome. ORFs are represented as block arrows. Arrows of paralogous genes have the same color with the exception of the grey-colored arrows, which represent unique genes. The names of the ORFs with predicted functions are indicated above the arrows, with ORFs of predicted membrane-targeted proteins indicated with *. The ORF numbers below the arrows correspond to annotations listed in Table 3 with # indicating genes that contain mutations separating them in two truncated ORFs. However, the *tra5* ORFs of PMU4 contains separate A and B ORFs that may produce a full-length transposase upon a single frameshifting event (53).

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396