*Short title:* *Genomic drafts of 'stolbur' phytoplasma from MDA templates*

# Generation and analysis of draft sequences of 'stolbur' phytoplasma from multiple displacement amplification templates

**Jelena Mitrović[1], Christin Siewert[2], Bojan Duduk[1], Jochen Hecht[3,4], Karin Mölling[4,5], Felix Broecker[4,6], Peter Beyerlein[7], Carmen Büttner[2], Assunta Bertaccini[8], and Michael Kube[2*]**

[1]*Laboratory of Applied Phytopathology, Institute of Pesticides and Environmental Protection, Banatska 31b, P.O. Box 163, 11080 Belgrade, Serbia*

[2]*Department of Crop and Animal Sciences, Humboldt-Universität zu Berlin, Lentzeallee 55/57, 14195 Berlin, Germany*

[3]*Berlin-Brandenburg Center for Regenerative Therapies, Augustenburger Platz 1, 13353 Berlin, Germany*

[4]*Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany*

[5]*University of Zurich, Gloriastr 32, 8006 Zurich, Switzerland*

[6]*Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany*

[7]*University of Applied Sciences, Wildau 15745, Germany*

[8]*DipSA, Plant Pathology,* Alma Mater Studiorum *University of Bologna, viale Fanin 42, 40127 Bologna, Italy*

[*]*Correspondence should be addressed to Michael Kube: Michael.Kube@agrar.hu-berlin.de*

Email-contacts:

Jelena Mitrovic, JelenaMitrovic@pesting.org.rs

Christin Siewert, Christin.siewert@agrar.hu-berlin.de

Bojan Duduk, Bojan.Duduk@pesting.org.rs

Jochen Hecht, Hecht@molgen.mpg.de

Karin Mölling, Moelling@imm.uzh.ch

Felix Broecker, Felix.Broecker@mpikg.mpg.de

Peter Beyerlein, Peter.Beyerlein@tfh-wildau.de

Carmen Büttner, Carmen.Buettner@agrar.hu-berlin.de

Assunta Bertaccini, Assunta.Bertaccini@unibo.it

Michael Kube, Michael.Kube@agrar.hu-berlin.de

**Abstract**

Phytoplasma associated diseases are reported for more than a thousand plant species worldwide. Only a few genome sequences are available in contrast to the economical importance of these bacterial pathogens.

A new strategy was used to retrieve phytoplasma strain-specific genome data. Multiple displacement amplification was performed on DNA obtained from less than three grams of plant tissue from tobacco and parsley samples infected with stolbur strains. Random hexamers and Phi29 polymerase were evaluated with and without supplementation by group-assigned oligonucleotides providing templates for Illumina's sequencing by synthesis approach.

Metagenomic drafts derived from individual and pooled strain-specific *de novo* assemblies were analysed. Supplementation of the Phi29 reaction resulted in an about 2-fold enrichment of the percentage of phytoplasma assigned reads and thereby improved assembly results. The obtained genomic drafts represent the largest data sets available from 'stolbur' phytoplasmas. Sequences of the two strains (558 kb, 448 proteins and 516 kb, 346 proteins, respectively) were annotated allowing the identification of prominent membrane proteins and reconstruction of core pathways. Analysis of a putative truncated sucrose phosphorylase provides hints on sugar degradation. Furthermore, it is shown that drafts obtained from repetitive-rich genomes allow only limited analysis on multi-copy regions and genome completeness.

**Key Words**: 'stolbur' phytoplasma, MDA, metagenome analysis, metabolism

## Introduction

Phytoplasmas are bacterial pathogens infecting plant and insects. They are associated with diseases in several hundred plant species including many important crops [McCoy et al., 1989]. Recently, more than 30 species have been assigned to the provisory taxon '*Candidatus* Phytoplasma' in the phylum *Tenericutes* [Kube et al., 2012]. The assignment as a candidate clade reflects the resistance of phytoplasmas to attempts of a cell-free cultivation in the past, which so far also slowed down genome research in phytoplasma; however a first cell-free cultivation of phytoplasma strains was recently reported for a few strains [Contaldo et al., 2012].

As a result of their evolutionary adaptation to intracellular life in plants and insects phytoplasmas have small genomes encoding a minimal metabolism [Kube et al., 2012; Oshima et al., 2004]. Only four complete genomes were determined so far encompassing '*Candidatus* Phytoplasma asteris' strains OY-M (853 kb) and AY-WB (707 kb), '*Ca*. P. australiense' (880 kb), and '*Ca*. P. mali' (602 kb) [Bai et al., 2006; Kube et al., 2008; Oshima et al., 2004; Tran-Nguyen et al., 2008]. Recently, the draft genomes of vaccinium witches' broom phytoplasma (647,754 nt in 272 contigs), Italian clover phyllody phytoplasma strain MA (597,245 nt in 197 contigs), poinsettia branch-inducing phytoplasma strain JR1 (631,440 nt in 185 contigs) and milkweed yellows phytoplasma (583,806 nt in 158 contigs) were published [Saccardo et al., 2012]. Phytoplasma-genome data were used for downstream experiments ranging from the design of PCR primers for marker genes and/or strain differentiation [Neriya et al., 2011; Seemüller et al., 2011] to the identification of effectors changing the plant phenotype [Hoshi et al., 2009; Sugio et al., 2011]. Genome sequences also allow the bioinformatical reconstruction of the major pathways in phytoplasmas [Kube et al., 2012].

This study deals with ‘stolbur’ phytoplasmas, which are infecting a wide range of cultivated and wild plants [Garnier, 2000]. This phytoplasma now classified as ‘*Ca*. P. solani’ [Quaglino et al., 2013], together with closely related ‘*Ca*. P. australiense’, belongs to 16S ribosomal group XII [IRPCM, 2004]. Some very important plant diseases and severe crop losses are associated with this phytoplasma such as ‘bois noir’ on grapevine, corn reddening and stolbur of solanaceous plants. ‘Stolbur’ disease was first described in Serbia starting in 1949 and since then it has been reported in many cultivated crops and ornamental plants as well as in wild plant species and its vectors. First insights on the genetic information encoded by ‘stolbur’ phytoplasma were provided by sequencing clones derived from amplified DNA enriched by double suppression subtractive hybridization (SSH) from *Catharanthus roseus* L. infected with strain PO. Cimerman and colleagues (2006) determined 181 sequences representing 113,209 nucleotides encoding 194 coding sequences of their estimated 820-850 kb sized genome. Despite this chromosome size estimation, loss of larger genome fragments during the propagation in infected plant hosts is described [Brown et al., 2007] and may also indicate the occurrence of smaller chromosomes in nature. The impact of horizontally transferred genetic material [Kube et al., 2012; Wei et al., 2008], transposon-mediated gene duplication [Bai et al., 2006] or group II intron [Cimerman et al., 2006] on these processes cannot be estimated for ‘stolbur’ phytoplasmas so far.

Since the major problem for studying phytoplasmas in its original hosts is the low concentration of pathogen in infected plants [Tran-Nguyen and Gibb, 2007] the enrichment of phytoplasma DNA or depletion of host genomic DNA is a critical step. Several approaches were used with pulsed field gel electrophoreses and gradient centrifugation being the most successful strategies applicable to many different types of starting material [Cimerman et al., 2006; Jiang et al., 1988; Seddas, 1993]. PFGE technique provides the highest DNA quality with respect to the purity [Tran-Nguyen and Gibb, 2007]. However, caesium chloride (CsCl) buoyant gradient centrifugation [Kollar et al., 1990] is one of the most commonly used

method to enrich phytoplasma DNA. Other methods such as suppression SSH e.g. in combination with mirror orientation selection result in a selected amplification of extrachromosomal DNA [Tran-Nguyen and Gibb, 2007] or for a doubled SSH in a biased representation of the chromosome [Yilmaz et al., 2010]. Isothermal amplification using Phi29 DNA polymerase was also used to obtain high quantities of phytoplasma DNA from low amount templates [Tran-Nguyen and Gibb, 2007] since this system mediates whole genome amplification (WGA) from single cells [Hosono et al., 2003; Spits et al., 2006]. It is characterised by its strand displacement activity and its impressive processivity in a highly accurate manner [Esteban et al., 1993]. Amplicons of more than 70,000 nt can be produced [Blanco et al., 1989]. The technical application of this enzyme was introduced by the rolling circle amplification (RCA) using random primers [Dean et al., 2001; Lizardi et al., 1998]. This approach was improved by the introduction of thiophosphate-modified random hexamers [Dean et al., 2002] and renamed multiple displacement amplification (MDA) taking into account that also linear products are amplified. Consequently, MDA was used in combination with target-specific primers for diagnostic purposes [James et al., 2011; Xu et al., 2008]. Here, we present a new MDA-based strategy to obtain DNA templates for deep sequencing, which was used to generate annotated genomic draft sequences of two ‘stolbur’ strains from naturally infected tobacco and parsley.

## Experimental Procedures

An overview of the workflow of the experiment carried out is given in figure 1 including key steps such as data generation, analysis of the metagenomic data, calculation of a potential enrichment by selected oligomers, analysis of the deduced protein content of the two strains, determination of the *sucP* transcript, comparison of the gene content of different strains and limits resulting from the sequencing of repeat-rich genomes in the assembly.

*DNA extraction and initial enrichment*

Two ‘stolbur’ strains from Northern Serbia were used: 284/09 originated from tobacco plant collected in Ečka in September 2009, grafted to tobacco seedling and maintained in micropropagation. On this sample Dellaporta enrichment [Dellaporta et al., 1983] was used followed by CTAB extraction [Ahrens and Seemüller, 1992]. Stolbur strain 231/09 originated from parsley plant (*Petroselinum sativum*) collected in Pančevo in October 2009 was the second one, and for it CTAB extraction and no initial enrichment were used. Half a gram of starting material was used for strain 231/09 and less than 3 g for 284/09.

*MDA control experiments*

The Illustra GenomiPhi DNA amplification kit V2 (GE Healthcare, Munich, Germany) was used for all Phi29 based amplifications in this study. Amplification and subsequent purification (Qiagen clean-up recommendations for repli-G products: www.qiagen.com) was performed according to the manufacturer’s instructions for a reaction volume of 20 µl without modifications (control experiments). One reaction for DNA templates obtained from parsley infected with 231/09 and one for tobacco infected with 284/09 was set up for comparison.

*Calculation of oligomers for the use of enrichment oligonucleotides*

Random hexamers and/or oligonucleotides corresponding to highly frequent oligomers encoded in the phytoplasma genomes were used for DNA amplification. Oligomers were calculated on a sequence concatamer consisting of the four complete chromosomes of ‘*Ca*. P. asteris’ strains AY-WB (CP000061) and OY-M (AP006628), ‘*Ca*. P. mali’ (CU469464) and ‘*Ca*. P. australiense’ (AM422018). The obtained phytoplasma concatamer was divided by a Perl (http://www.perl.org/) script into 102 windows with a size of about 30 kb. All possible 21mers in each respective window were calculated and counted using a sliding window

approach. Selected oligomers correspond to the used oligonucleotides within the next steps. Oligomer sequence and its appending count of the respective window were saved with the help of a hash during processing and sorted by their count. To reduce calculation time and storage space only the 100,000 most frequently oligomers for each window were saved. Oligomers sequences were excluded, if they do not contain two bases of cytosine, two bases of guanine or one base of guanine and one base of cytosine at least. This step was performed to increase the specificity. Most frequently encoded oligomers were determined and a minimal set of 28 oligomers (21 bases in length each) was selected covering 95% of the concatamer, if product size of 30 kb is assumed. The set of enrichment oligonucleotides (Eurofins MWG Synthesis GmbH, Ebersberg, Germany) consists of:

5’-GATAATTTCATATCTTTTCAA-3’, 5’-TTTTAAAAATTCTTTAAGACT-3’,
5’-TTTTTTGGAAAAAAATGTTGT-3’, 5’-ATTCAATTAAATTTAATAAAC-3’,
5’-ACACTTAGATGCTTTCAGTGT-3’, 5’-CTGTTTCATTTATTTAGACAA-3’,
5’-TTTGAAAAGATATGAAATTAT-3’, 5’-TATTTATATCGATCTTTTTTT-3’,
5’-AAAGATATGGAAATATCTAAT-3’, 5’-TTTTCATTTGGTTCATTTTTT-3’,
5’-GGACATTTAATTTCTAATAAA-3’, 5’-TATTAAAAACTAAAATATCGA-3’,
5’-TACAAACGTGATACAATGGCT-3’, 5’-TTAAAATGGCGAGAAATTACA-3’,
5’-AATTTTTAAATCAAAAAATGT-3’, 5’-GAAATGCTTAAATTAGCTGAT-3’,
5’-TTATCATGAAGCAGGACATGC-3’, 5’-TATTGATAAATTTTATAAGGA-3’,
5’-TGACTCCTCAATTAGAAGAAC-3’, 5’-CGTTTGAGAAGGGATTCGAAC-3’,
5’-ACTAAATTGTTTATTAAAAAA-3’, 5’-GTTCGAATCCCTTCTCAAACG-3’,
5’-AAAATCATATTGATATAATAA-3’, 5’-ACGTAATGTTATTTTAGAAAA-3’,
5’-TGTCTAAGAAAAAAATTATTT-3’, 5’-ATTATATCAATATGATTTTTA-3’,
5’-GTTAAAAATTAACCATTTATT-3’, 5’-AATCAAAAAAATAATAAGGAG-3’.

*MDA using phytoplasma enrichment oligonucleotides*

Amplification and subsequent purification was performed according to the manufacturer's instructions with some modifications. Phi29 reaction volumes were supplemented by additional oligonucleotides increased by 1 µl to a final reaction volume of 21 µl. The set of enrichment oligonucleotides (100 pmol each) was added to the reaction mix including the random hexamers. The set of enrichment oligonucleotides is present in a final concentration of 0.4 µM. Experiments were performed for both strains.

*MDA using group-specific oligonucleotides*

In an independent control experiment P1/P7 primers [Deng and Hiruki, 1991; Smart et al., 1996] were used in different concentrations in combination with random hexamers. The three experiments contain P1/P7 primer with a final concentration of 0.4 µM, 3.3 µM and 5.5 µM. Experiments were carried out on tobacco sample infected with strain 284/09.

The experiment with the highest primer supplement were carried out using primers and DNA template vacuum dried in a speedvac (Savant, Farmingdale, NY, U.S.A.) and resuspended in the sample buffer of the illustra GenomiPhi DNA amplification kit V2 (GE Healthcare, Munich, Germany) for 30 minutes to keep equal reaction volumes.

*Sequencing by synthesis*

About 3-5 µg of each MDA product was used for library preparation from Illumina sequencing. Barcoded libraries were prepared using the NEBnext DNA sample prep Kit (New England Biolabs, Ipswich, Massachusetts, U.S.A.) according to manufacturer's instructions. Libraries were sequenced on a Genome Analyzer IIx (Illumina, San Diego, California, U.S.A.) in a 36 base single read multiplex run.

De novo *assemblies*

Each sample was assembled using the CLC Genomics workbench V4.5 (http://www.clcbio.com/). Standard parameter for read trimming and *de novo* assembly was used with one exception. The similarity value was increased to 0.9. The average used read length of the quality trimmed reads was 35 bases. The minimal size for contiguous sequences (contigs) was set to 300 bp. In addition, all obtained data of each strain were assembled and used for individual draft genome analysis in the following steps.

*Taxonomical assignment*

Contigs were compared via BLAST [Altschul et al., 1997] against NRPROT (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). BLASTX results were uploaded in MEGAN (MEta Genome Analyzer) [Huson et al., 2011]. Contigs were handled as reads in MEGAN using a coverage level of one. All sequences with an assignment to the phylum *Tenericutes* were selected for automated annotation [Rabus et al., 2002].

*Calculation of enrichment*

Enrichment was evaluated by read counts in two different approaches. First, the percentage of reads showing perfect matches (100% identity) on the chromosome sequence of the four complete genomes (AP006628.2, CP000061, CU469464, CU469464) was calculated. This was performed by megablast using a word length of 20 and subsequently filtering of the output by a Perl script. Second, the percentage of phytoplasma assigned reads building contigs was calculated from the assembly.

*Draft genome annotation*

Contigs for each strain were obtained by the assembly of all read data and subsequent taxonomical selection as described above. Contigs were merged in a draft sequence but kept

separated by 100 Ns. The two drafts were analysed by the automated annotation pipeline HTGA [Rabus et al., 2002] and manually curated in Artemis [Rutherford et al., 2000]. Annotated draft sequences were deposited in Genbank (strain 284/09 Acc. no. FO393427 and strain 231/09 Acc. no. FO393428).

*Expression of* sucP *gene in strain 284/09*

Total RNA was prepared from tissue-cultured tobacco infected by 'stolbur' strain 284/09 using nucleoSpin RNA II extraction kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions and using a final elution volume of 50 µl. Eight microliters of RNA was treated with one unit of DNase I (Thermo Scientific, Bremen, Germany). Aliquots of 4.5 µl DNase treated RNA were used as template for amplification in RT-PCR and PCR reactions to examine expression of *sucP* and absence of DNA, respectively. Primers sucP-left (5'- AGTTTTTGCACCAGGCATTC -3') and sucP-right (5'- GTGTGGGGGTTGTAATTTCG -3') were designed for amplification of partial *sucP* sequence (expected product of 242 bp). RT-PCR assay with the sucP-left/right primer pair was carried out on DNase treated RNA in 25 µL-scale RT-PCR reaction mix containing 4.5 µL template RNA, 1X OneStep RT-PCR buffer, 0.4 mM dNTP mix, 0.4 µM of each primer and 1µl OneStep RT-PCR enzyme mix (Qiagen, Hildesheim, Germany).

PCR assays were carried out on DNase treated RNA and total DNA in a 25 µL-scale PCR reaction mix containing 4.5 µL template RNA or 1 µL template DNA, 1× PCRMasterMix (Thermo Scientific, Bremen, Germany) and 0.4 µM of each primer. PCR negative control reaction was set up without DNA.

Reverse transcription reaction was performed for 30 min at 50$^{o}$C followed by 13 min initial denaturation at 95$^{o}$C. Subsequent PCR assay contains a denaturation step of 2 min at 94$^{o}$C and a three step PCR cycling (94$^{o}$C, 30 sec; 57$^{o}$C, 30 sec; 72$^{o}$C, 1 min) repeated 35 times and a final extension step of 5 min for 72$^{o}$C. Amplification or absence and size of PCR products

were analyzed by agarose gel electrophoresis using 6 µl of RT-PCR or PCR products. 100 bp and 1 kb DNA ladder were used for size comparison (Thermo Scientific, Bremen, Germany).

*Comparative analysis of 'stolbur' strains*

The two 'stolbur' draft sequences from Serbia were compared by BLASTN [Altschul et al., 1997] using strain 284/09 as reference. Calculation was performed using a word length of 300 and low complexity filter off. BLASTN-output was filtered for a minimal sequence identity of 99% and average alignment length was calculated by a Perl script. Conservation of deduced protein content was estimated using the draft sequence of strain 284/09 as subject and the deduced proteins of strain 231/09 as query in a TBLASTN comparison. In a second approach, deduced proteins of the Serbian stolbur strains were compared via BLASTP. BLAST-outputs were filtered for identity of 99% of the aligned regions by MSPcrunch [Sonnhammer and Durbin, 1994]. Comparison on obtained data versus published 'stolbur' phytoplasma determined in the study of Cimerman *et al*. (2006) was performed by BLASTP [Altschul et al., 1997] and filtering the output by MSPcrunch [Sonnhammer and Durbin, 1994].

*Limits of short read assemblies*

The problems of the assembly of repeat-rich genomes resulting from short reads were simulated by an *in silico* approach. Therefore, the chromosome sequences of '*Ca*. P. australiense' (acc. no. AM422018) and *Escherichia coli* strain K12 substr. MG1655 (acc. no. U00096.2) were used as template for the ART simulation tool [Huang et al., 2012] to obtain simulated Illumina read data sets. Single read data sets were calculated with a length of 35 b, 100 b and paired-end reads with (2x) 100 b covering the chromosomes 168-fold each. Reads were assembled (see above), total contig length (TCL) was calculated and contigs extracted.

*Mapping on the chromosome of* 'Ca. *P. australiense*'

Chromosome information on protein encoding sequences and potential pseudogenes of '*Ca*. P. australiense' were extracted from the genome (acc. no. AM422018). Potential multi-copy genes were identified by BLASTClust (http://www.ncbi.nlm.nih.gov) also considering the annotated pseudogenes, which are so far not experimentally validated. '*Ca*. P. australiense' contigs obtained from the assembly of simulated Illumina reads (ART approach see above) were assigned to corresponding positions on its chromosome by BLASTN using a wordsize of 100 b and subsequent filtering for a minimal identity of 99% by MSPcrunch. Deduced protein coding sequences of the two Serbian 'stolbur' strains were assigned to potential homologs of '*Ca.* P. australiense' by BLASTP (e-value cut-off: 1e-15 and MSPcrunch identity filter cut-off: 30%). First hit was considered in both analyses, if multiple hits show an identical BLAST score. GC plot calculation and visualization of results was performed in Artemis.

## Results and discussion

*Obtained sequence data*

Sequencing by synthesis (SBS) was performed for nine barcoded libraries (one lane of the flow-cell) resulting in 20,588,492 quality passed short-reads with an average length of 35 bases. Obtained read numbers/sample vary from 2,290,270-3,494,601 reads (**Table 1**). In total, above 741 Mb of quality filter-passed sequence information was generated. Experiments limited to Dellaporta enrichment approach and Phi29 amplification using random hexamers for strain 284/09 and Phi29 amplification using random hexamers alone for strain 231/09 resulted in short total contig length (TCL) of 211.6 kb and 7.6 kb, respectively. Higher TCL was obtained in all Phi29 reactions supplemented by oligonucleotides (P1/P7 or

316 oligonucleotides based on calculated oligomers). Furthermore, a lower G + C content was
317 observed for the contigs derived from these experiments.
318 Percentages of reads assigned to phytoplasma containing contigs were calculated and an
319 enrichment was observed **(Table 2)**. Supplementation by the calculated oligomers resulted in
320 a 2.3 and 2.4-fold increase of assigned phytoplasma reads and in a 2.2 and 5.9-fold increase of
321 the percentage of the phytoplasma assigned TCL for strain 284/09 and strain 231/09
322 respectively. A TCL of 460 and 439 kb was reached in these experiments (**Table 1**). Long
323 TCLs resulted from the high percentage (80-81%) of phytoplasma assigned reads allowing a
324 contig formation (**Table 2**).
325 The highest P1/P7 oligonucleotide concentration did not result in an enrichment increase
326 (**Table 1 & 2**). This may result from insufficient resuspension of the previously dried
327 oligonucleotides. Low P1/P7 oligonucleotide supplementation also shows no significant effect
328 compared to the usage of the hexamers without additional supplementation. These results can
329 be observed taking into account the increase of reads in contigs assigned to phytoplasma but
330 also by an independent calculation based on a simplified mapping of reads from this study on
331 the four complete phytoplasma genomes.
332
333 *Strain data assemblies*
334 Data obtained for each strain were pooled and assembled. The unselected metagenomic draft
335 sequence of the strain 284/09 plant sample reaches a length of 1,340 kb while 915 kb
336 sequence was obtained for the strain 231/09 infected sample. Draft sequences were organized
337 in 685 contigs for strain 284/09 and 739 contigs for strain 231/09 (minimal contig size 300
338 bp). The contig numbers result in 1,965 bp and 1,238 bp average contig length, respectively.
339 About 19% (128/685) of the contigs of strain 284/09 and 40% (298/739) of strain 231/09
340 could be assigned to phytoplasma **(Table 3)** using the MEGAN approach [Huson et al.,
341 2011]. Below 1% of the initial contigs (assignment to the phylum *Tenericutes*) could not be

assigned to '*Ca*. Phytoplasma' and were removed manually during the gene annotation. The average contig length of the phytoplasma assigned sequences is increased with 4,356 bp (168-fold sequencing coverage) and 1,731 bp (34-fold sequencing coverage) compared to unselected contigs and results from a lower plant background. The draft sequences show an average identity of 99% to each other in aligned regions (above 457 kb).

The 448 (strain 284/09) and 346 (strain 231/09) protein coding genes were assigned as completely determined. The shorter contig length in strain 231/09 results in a lower number of complete genes and a higher number of partial genes. In addition, differences in the completeness and quality are indicated by the number of 27 and 8 tRNAs encoded in the two strains. However, successful application of this genomic draft approach can be estimated by the functional assignment made on 328 deduced complete proteins for strain 284/09 and 238 for lower covered strain 231/09. 452 of 573 deduced protein sequences of strain 231/09 are also encoded in strain 284/09 showing an identity of above 99%.

A high portion of the 520 deduced proteins of 'stolbur' strain 284/09 can be also identified in the protein data set of the completely determined chromosome of '*Ca*. P. australiense'. However, only one mapped protein of strain 284/09 shows 99% identity to '*Ca*. P. australiense' but 438 at least 80% identity. The data set from 'stolbur' PO generated by SSH approach produced 194 deduced protein sequences [Cimerman et al., 2006]. 186 potential orthologs were identified in strain 284/09 showing an identity of at least 30%; 80 of the identified proteins show an identity of 99%.

*Analysis of the genomic core of 'stolbur' strains 284/09 and 231/09*

Complete gene sets encoding proteins for replication, DNA modification and structure and DNA repair were identified in both genomes [Kube et al., 2012]. Both strains encode the excision repair complex *uvrABC* comparable to the one of '*Ca*. P. asteris' strains OY-M and AY-WB and '*Ca*. P. australiense', this latter belongs to the same phylogenetic branch. Results

support the separated evolutionary genome condensation of '*Ca.* P. mali' that is lacking these genes.

Furthermore, the gene content necessary for transcription was identified in both draft sequences. Differences in the gene content due to the level of completeness of drafts are visible in the gene sets involved in translation, e.g. for ribosomal genes. The draft of strain 284/09 contains the complete set of ribosomal genes, while the draft of strain 231/09 lacks six genes (*rplJ, rpmF, rpmH, rpmI, rpsL* and *rpsN*). It was possible to identify all tRNA synthetases in the draft of strain 284/09, while it was impossible to identify the genes for prolyl-tRNA and lysidine-tRNA synthetase in strain 231/09.

The conserved content of membrane-associated genes was identified in both draft sequences including the associated gene set assigned to lipid synthesis [Kube et al., 2012]. Common ABC-transporters for the ATP-dependent putative import of manganese/zinc, cobalt, spermidine/putrescine, sugars or putative glycerol-3-phosphate, oligopeptides and methionine are encoded in the sequencing obtained. Furthermore, a putative thiamine transporter of the ThiA/YuaJ family was identified. Both strains also share the content of multidrug efflux pumps (MdlA/B, NorM), P-type ATPases and the large-conductance mechanosensitive channel protein with the other sequenced phytoplasmas.

The prominent membrane proteins Vmp1 and Stamp (Amp) of 'stolbur' were identified in both draft sequences. Both proteins are supposed to be involved in phytoplasma-host interaction [Cimerman et al., 2009; Fabre et al., 2011]. StamP is supposed to show a conserved sequence synteny with the order *groEL-amp-nadE* in several phytoplasma genomes [Fabre et al., 2011] but not in '*Ca*. P. mali' (acc. no. CU469464). This synteny was also identified in the two strains examined in this study. This is in agreement with the identity of 91% reached in the alignment of strain 284/09 and strain 231/09 of StamP and VmP1 from stolbur phytoplasma [Fabre et al., 2011].

In both ‘stolbur’ strains, gene sets necessary to build the general Sec-dependent pathway in phytoplasma were identified (*ffh*, *ftsY*, *secA*, *secE*, *secY*) except for *yidC*, which was not identified in strain 231/09.

The energy metabolism encode the Embden-Meyerhof-Parnas pathway [Oshima et al., 2004] and the suggested alternative pathway from malate (or a similar substrate) to acetate [Kube et al., 2012]. In consequence, the genes encoding the upper part of the glycolysis were identified (phosphoglucose isomerase, Pgi; phosphofructo-kinase, PfkA; fructose-biphosphate aldolase, Fba; triosephosphate isomerase, TpiA). In contrast to ‘*Ca*. P. mali’ but in agreement with the other fully sequenced phytoplasmas, the genes encoding proteins involved in the energy yielding part were also identified (glyceralaldehyde-phosphate dehydrogenase, GapA; phosphoglycerate mutase, Pgm; enolase, Eno; pyruvate kinase, PykF). The suggested alternative energy-yielding pathway encoded in all four completely determined genomes of phytoplasma strains is also encoded by ‘stolbur’ strains from Serbia [Kube et al., 2012]. It encodes the steps of malate uptake (symporter MleP), oxidative decarboxylation (SfcA), generation of Acetyl-CoA by the pyruvate dehydrogenase multienzyme complex (AcoAB, AceF, Lpd), formation of Acetyl-phosphate by the suggested phosphotransferase (PduL-like protein) and acetate (AckA) including the gain of one ATP [Kube et al., 2012]. In contrast to this second pathway, glycolysis is dependent on the supply of a phosphorylated hexose at the beginning. The option of a direct uptake of trehalose-6-phosphate or sucrose-6-phosphate and subsequent processing with respect to the particular environment of phytoplasmas has to be taken into consideration [Kube et al., 2012]. Indeed, no hints for an encoded membrane bound phosphoenolpyruvate-dependent phosphotransferase system (PTS) were provided for phytoplasmas so far and it remains still unclear how phytoplasmas perform this step. A phosphoenolpyruvate-dependent PTS activity was shown for *Mycoplasma gallisepticum*, *M. mycoides* subsp. *mycoides* and *M. mycoides* subsp. *capri* but not for *Acholeplasma laidlawii*

418 [Cirillo and Razin, 1973]. The last finding might be remarkable because *Acholeplasma* are
419 paraphyletic to the genus '*Ca*. Phytoplasma' [Gundersen et al., 1994].
420 However, both 'stolbur' strains encode a truncated sucrose phosphorylase (SucP). Sequence
421 comparison of *sucP* and flanking sequences show 100% identity on nucleotide sequence.
422 Sucrose may be taken up by some phytoplasmas from the phloem sap via a sugar ABC-
423 transporter [Kube et al., 2012] or may be a permease [Reid and Abratt, 2005]. The *sucP* gene
424 is encoded between a Zinc-dependent protease (TldD) and a multidrug efflux pump (NorM)
425 in both strains. SucP mediates the generation of alpha-D-glucose-1-phosphate and beta-D-
426 fructofuranose from sucrose and to a lesser extent hydrolysis glucose-1-phosphate [van den
427 Broek et al., 2004]. While TldD and NorM protein show highest similarities to phytoplasmas,
428 SucP shows its lowest e-values in BLASTP against NRPROT to *Leuconostoc mesenteroides*
429 (e-value $6e^{-87}$; acc. no. ABW71903) followed by members of the genus *Lactobacillus* (e-value
430 $6e^{-84}$; refseq acc. no. NC_009513.1). Proteins show a high identity of around 70% to the SucP
431 of strain 284/09. The deduced protein length of 485-491 aa significantly differs from the SucP
432 of the two Serbian 'stolbur' strains (136 aa). SucP is also encoded in '*Ca*. P. australiense'
433 (annotated as GtfA) and '*Ca*. P. asteris' strain OY-M. While it might be also truncated in
434 strain OY-M, '*Ca*. P. australiense' encodes the putative full-length protein. It was impossible
435 to detect this protein in '*Ca*. P. asteris' strain AY-WB or in'*Ca*. P. mali' genomes. The coding
436 and truncation of *sucP* in phytoplasmas is remarkable because sucrose represents the most
437 prominent sugar in the phloem sap [Vanhelden et al., 1994]. Surprisingly, expression of *sucP*
438 is confirmed for strain 284/09 by RT-PCR (**Figure 2**). One may speculate, if the truncated
439 *sucP* lost its original function. However, the subsequent processing remains unclear because
440 alpha-D-glucose-1-phosphate has to be converted to alpha-D-glucose-6-phosphate by the
441 phosphoglucomutase (PgmA). Such an alpha-D-phosphohexomutase is encoded in
442 *Acholeplasma laidlawii* strain PG-8A (acc. no. YP_001620839) and in several mycoplasmas

443 but is not identified in phytoplasmas so far. This might also be due to the weak genomic
444 database for phytoplasmas.
445 This little shunt that would allow phytoplasmas to (I) metabolize the most prominent sugar in
446 phloem sap, (II) overcome the lack of a PTS system and (III) provide a needed entry enzyme
447 of the upper part of the glycolysis. It is likely that the ancestors of the phytoplasmas were able
448 to use this shunt for the degradation of sucrose or a similar disaccharide. Furthermore, it
449 raises the question, if the other prominent disaccharide trehalose of the insect vectors was
450 utilized in a similar manner by phytoplasmas. Phytopathogenic spiroplasmas perform such a
451 switch from glucose and fructose in the plant to trehalose in the hemolymph of the leafhopper
452 vector [Gasparich, 2010; Gaurivaud et al., 2000].
453
454 *Reached genome coverage using short reads for assembly*
455 No estimation is made on the completeness of the draft genomes of the examined 'stolbur'
456 strains. The complete genome of the related '*Ca*. P. australiense' is heavily characterized by
457 repeats covering 24% of the sequence [Tran-Nguyen et al., 2008]. There is a risk of
458 misassemblies due to the shared high similarity within the individual repeat groups. Conflicts
459 occurring during the assembly of these regions result in rejection of reads from the assembly.
460 It is likely that several repetitive elements are encoded in both strains. Identification of
461 transposases (3 proteins in strain 284/09 and 1 in strain 231/09) and a group II intron reverse
462 replicase (5 proteins in strain 284/09 and 3 in strain 231/09) support such an estimation for
463 both samples.
464 A simulation based on the chromosome sequence of '*Ca*. P. australiense' with a length of
465 879,959 bp was performed calculating Illumina read sets with a size of 35, 100 and paired
466 reads of 2x 100 nucleotides using ART simulation tools [Huang et al., 2012]. None of the
467 performed *de novo* assemblies using the simulated read sets reached chromosome length
468 (single read length 35 b: 586 kb; single read length 100 b: 596 kb; paired-end reads length

100 b each: 616 kb). Contigs with a size of at least 300 b cover 67-70% of the chromosome. Misassemblies and conflicts resulting from repeats have a major impact on results. This is supported by a mapping of the 586 kb reached by the assembly of simulated reads from '*Ca*. P. australiense' on the chromosome in comparison to the location of potential pseudogenes and multi-copy genes (**Figure 3**). However, simulated assemblies show a sequence identity of 99% to the reference chromosome providing an estimation of the quality and usability of sequences of such a genomic draft. There is a high probability that also the obtained assemblies of the Serbian 'stolbur' strains show a similar problematic taking into account the distribution of the deduced proteins by mapping on '*Ca*. P. australiense'.

In comparison, assemblies based on read-simulation performed on the comparable repeat-poor 4,640 kb large *E. coli* strain K12 substr. MG1655 chromosome resulted in a similar chromosome length covering 97-98% considering contigs with a minimal length of 300 bp (single read length 35 b, TCL 4,518 kb; single read length 100 b: 4,532 kb; paired-end reads length 100 b each: 4,543 kb).

In conclusion, a direct estimation on completeness of the genomic draft of phytoplasmas appears to be risky: sequence quality of assembled regions is expected to show high confidence but usage of potential multi-copy genes for on-going studies directly from draft sequences is limited and needs additional evaluation.

*Other phytoplasma draft approaches*

Starting on a DNA template resulting from isothermal amplification, a short read sequencing approach was successfully used for the determination of two 'stolbur' draft genome sequences. It is shown that it is possible to generate phytoplasma genomic draft sequence from below a gram of starting material using low costs sequencing chemicals and comparable little hand on time. The presented first results of enrichment by group-assigned primers of target DNA in a metagenomic sample are basic for other studies. No plasmids were detected

in this study but it is likely that plasmid-containing templates are risky for such an approach [Tran-Nguyen and Gibb, 2007].

A higher number of premade assumptions are made by Firrao and colleagues providing four genomic draft sequences of phytoplasmas in a recently published study [Saccardo et al., 2012]. Major differences in the initial steps are the usage of a higher amount of starting material (10 µg of DNA) and no amplification prior to sequencing library construction. Finally assigned phytoplasma data were obtained by a coverage based selection of contigs and reads respectively, negative selection against a non-infected host plant data set and screening of contigs resulting from the selected read assembly against phytoplasma sequences in public databases. Phytoplasma drafts derived from the colonized model plant *Catharanthus roseus* reached TCL from 584-648 kb resulting from contigs with an average length of 2.3-3.7 kb of these samples. The draft sequence of an Italian strain of clover phyllody phytoplasma showing a higher level of colonization in *C. roseus* reached the highest percentage of assigned phytoplasma reads with 12% while the other samples reached only 3-6%. In comparison, none of the 'stolbur' templates reached more than 2.1% of assigned reads in our study and the non-amplified templates reached 0.8-0.9 % (**Table 2**). However, the higher sequencing coverage for 'stolbur' strain 284/09 (168-fold) resulted in an expected higher average contig size 4.4 kb. The strain 231/09 shows a decreased average contig length with 1.7 kb corresponding to 34-fold sequencing coverage and the usage of short reads (36 b) compared to the 100 b reads used in the study of Firrao and colleagues [Saccardo et al., 2012]. A high sequencing coverage for on-going *de novo* sequencing projects has to be anticipated in general because a shorter average contig length drastically influences the number of partial gene sequences in a draft (**Table 3**) and thereby the overall quality of the genomic draft.

**Acknowledgements**

520 COST-FA0807- STSM- 180111-006678, the German Research Foundation (DFG) through
521 projects KU 2679/2-1 and BU 890/21-1, the German Academic Exchange Service (DAAD)
522 through project 56266384 and the Ministry of Education and Science, Republic of Serbia,
523 grant number TR31043, supported this work.

**Table 1. SBS and assembly results (data assigned to phytoplasma strains is indicated by Ph+).**

| Template, experiment number | MDA-add-ons | Total no. of reads | No. of contigs | | G + C content | | Reads in contigs | | TCL of contigs (bp) | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Strain 284/09** | | | | **Ph+** | | **Ph+** | | **Ph+** | | **Ph+** |
| **624** | - | 2,970,584 | 233 | 112 | 42% | 38% | 581,102 | 112’122 | 630,092 | 211,581 |
| **621** | oligomers | 2,766,451 | 187 | 112 | 32% | 29% | 492,386 | 400’613 | 620,682 | 459,862 |
| **629** | P1/P7 (0.4 µM) | 2,786,556 | 165 | 41 | 41% | 29% | 266,109 | 43’171 | 410,496 | 81,915 |
| **628** | P1/P7 (3.3 µM) | 2,614,502 | 253 | 118 | 36% | 29% | 531,162 | 319’790 | 701,890 | 381,294 |
| **627** | P1/P7 (5.5 µM) | 3,494,601 | 244 | 106 | 31% | 29% | 711,084 | 428’407 | 896,441 | 497,771 |
| **Strain 231/09** | | | | | | | | | | |
| **626** | - | 2,290,270 | 30 | 6 | 39% | 32% | 35,682 | 5’357 | 57,681 | 7,630 |
| **623** | oligomers | 2,816,277 | 188 | 149 | 30% | 29% | 559,629 | 447’173 | 568,543 | 438,799 |

**Table 2. Percentages of phytoplasma assigned data (Ph+) in experiments.**

| Template, experiment no. | MDA-add-ons | % of Ph+ reads based on the total no. of reads in contigs | % of complete phytoplasma genome assigned reads calculated on total reads | % of Ph+ TCL based on total TCL |
|---|---|---|---|---|
| **Strain 284/09 (from tobacco)** | | | | |
| **624** | - | 19.3% | 0.9% | 33.6% |
| **621** | oligomers | 81.4% | 2.1% | 74.1% |
| **629** | P1/P7 (0.4 µM) | 16.2% | 0.9% | 20.0% |
| **628** | P1/P7 (3.3 µM) | 60.2% | 1.9% | 54.3% |
| **627** | P1/P7 (5.5 µM) | 60.2% | 1.5% | 55.5% |
| **Strain 231/09 (from parsley)** | | | | |
| **626** | - | 15.0% | 0.8% | 13.2% |
| **623** | oligomers | 79.9% | 1.9% | 77.2% |

532 **Table 3. Selection of 'stolbur' draft sequences and general genome features.**
533

| **Strain** | **284/09** | **231/09** |
|---|---|---|
| **Total no. of SBS reads** | 14,632,694 | 5,955,768 |
| **All contigs** | | |
| Number of contigs (>300 bases) | 685 | 739 |
| Average contig size (>300 bases) | 1,965 | 1,238 |
| TCL (>300 bases) | 1,346,432 | 915,060 |
| **Phytoplasma assigned contigs** | | |
| Number of contigs (>300 bases) | 128 | 298 |
| Average contig size (>300 bases) | 4,356 | 1,731 |
| TCL (>300 bases) | **557,538** | **515,758** |
| Number of implemented reads | 1,841,069 | 471,680 |
| Sequencing coverage | 168.2-fold | 34.3-fold |
| G + C content | 28.2% | 28.6% |
| CDS/ deduced proteins | 448 | 346 |
| additional partial CDSs | 72 | 227 |
| G + C content of CDSs | 29,6% | 29.8% |
| tRNAs | 27 | 8 |

534
535

**Figure 1. Overview of data generation.** Running numbers of the experiments are given in brackets.

**Figure 2. RT-PCR validating the expression of *sucP* gene in strain 284/09.** Lanes on the gel image: 1, 100 bp DNA ladder (M1); 2, RT-PCR product from DNA-free RNA sample (RT-PCR); 3, absent PCR product from DNA-free RNA sample (PCR1); 4, PCR control on genomic DNA (PCR2); 5, PCR negative-control (PCR3); 6, 1kb DNA ladder (M2). Arrows indicate the size of bands from the DNA standard.

**Figure 3. General chromosome information of '*Ca*. P. australiense' and visualization of the mapping results.** Circular patterns (from outside to inside): 1 (black circle), scale in base pairs of the chromosome; 2 & 3 (dark blue), annotated coding sequences on forward and reverse strand (AM422018); 4 (light blue), annotated pseudogenes (AM422018); 5 (red), multi copy genes; 6 (black), contigs obtained from simulated reads calculated from '*Ca*. P. australiense' chromosome sequence; 7 (green), assigned coding sequences of stolbur strain 284/09; 8 (gold), assigned coding sequences of stolbur strain 231/09; and 9 (olive and pink), G+C plot.

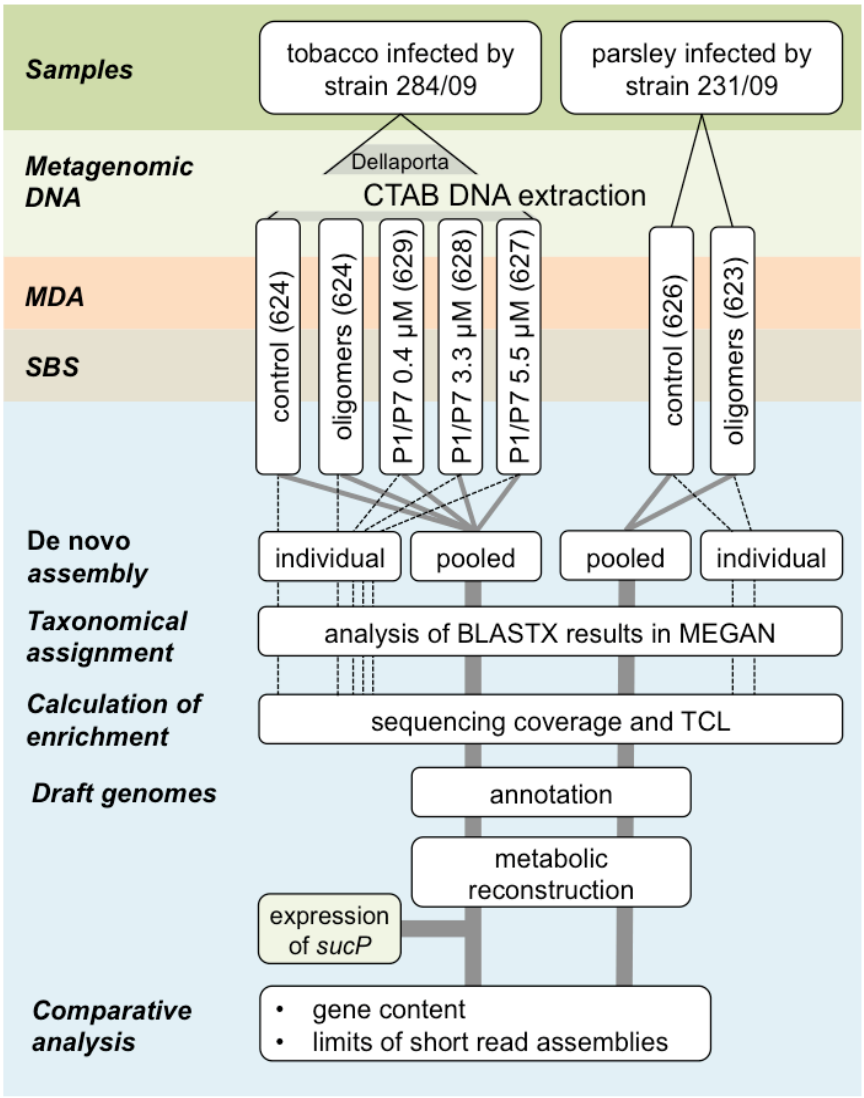
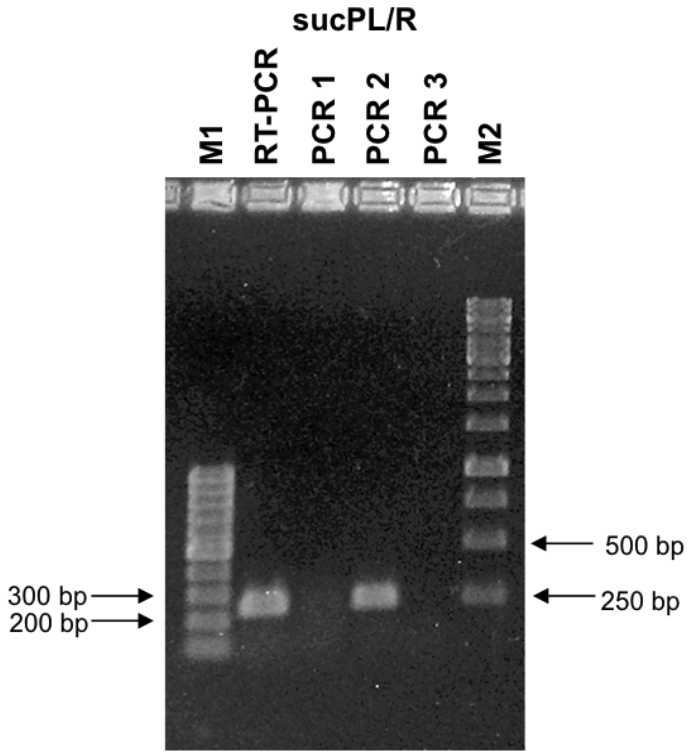

Samples
tobacco infected by strain 284/09
parsley infected by strain 231/09
Metagenomic DNA
Dellaporta
CTAB DNA extraction
MDA
SBS
control (624)
oligomers (624)
P1/P7 0.4 µM (629)
P1/P7 3.3 µM (628)
P1/P7 5.5 µM (627)
control (626)
oligomers (623)
De novo assembly
individual
pooled
pooled
individual
Taxonomical assignment
analysis of BLASTX results in MEGAN
Calculation of enrichment
sequencing coverage and TCL
Draft genomes
annotation
metabolic reconstruction
expression of sucP
Comparative analysis
• gene content
• limits of short read assemblies

**Figure 1.**

559




560

561 **Figure 2.**

562

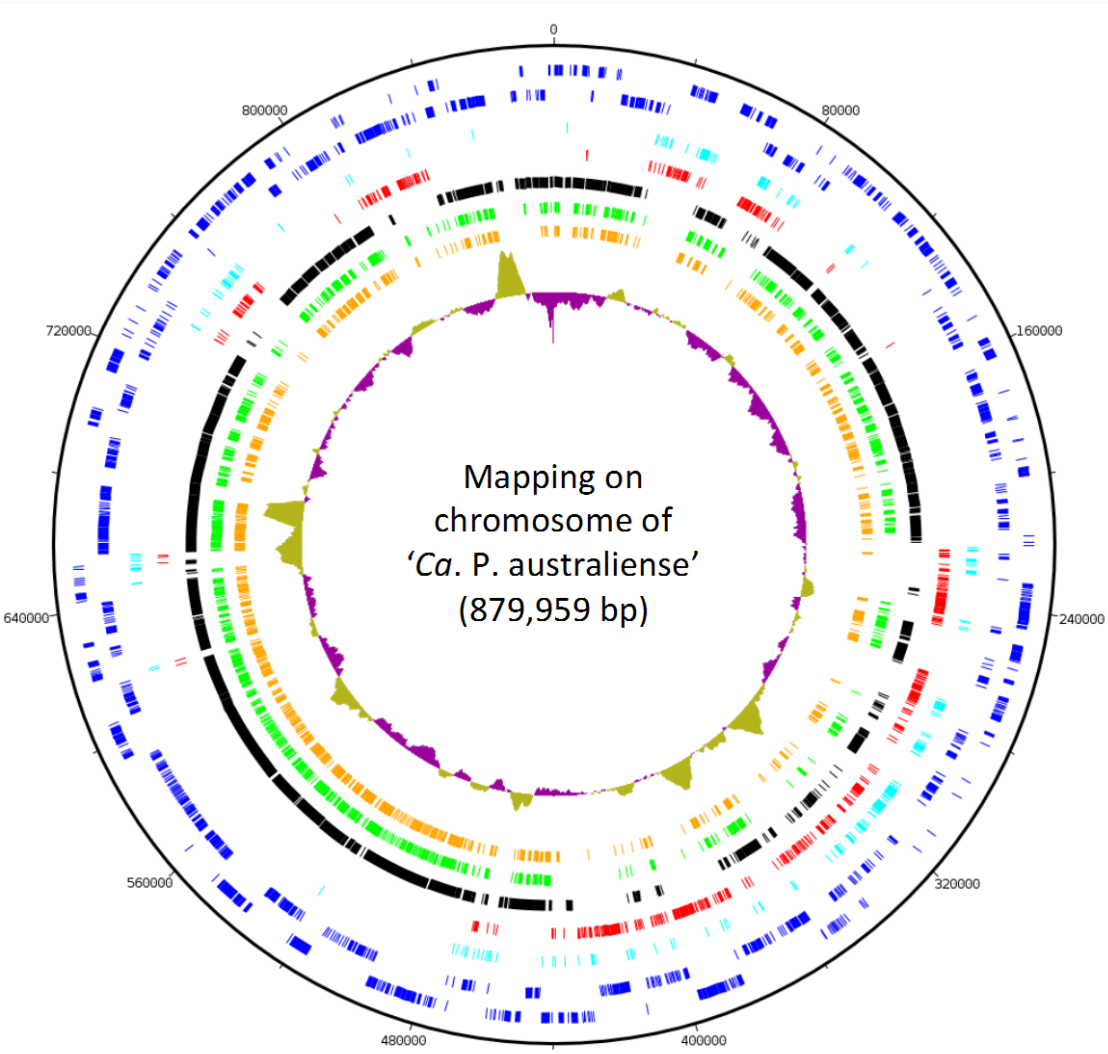

0
80000
160000
240000
320000
400000
480000
560000
640000
720000
800000
Mapping on
chromosome of
'Ca. P. australiense'
(879,959 bp)

**Figure 3.**

## References


Ahrens U, Seemüller E: Detection of DNA of plant pathogenic mycoplasmalike organisms by a polymerase chain reaction that amplifies a sequence of the 16S rRNA gene. Phytopathology 1992;82:5.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped blast and psi-blast: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-3402.

Bai X, Zhang J, Ewing A, Miller SA, Jancso Radek A, Shevchenko DV, Tsukerman K, Walunas T, Lapidus A, Campbell JW, Hogenhout SA: Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J Bacteriol 2006;188:3682-3696.

Blanco L, Bernad A, Lazaro JM, Martin G, Garmendia C, Salas M: Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem 1989;264:8935-8940.

Brown DR, Whitcomb RF, Bradbury JM: Revised minimal standards for description of new species of the class Mollicutes (division Tenericutes). Int J Syst Evol Microbiol 2007;57:2703-2719.

Cimerman A, Arnaud G, Foissac X: Stolbur phytoplasma genome survey achieved using a suppression subtractive hybridization approach with high specificity. Appl Environ Microbiol 2006;72:3274-3283.

Cimerman A, Pacifico D, Salar P, Marzachi C, Foissac X: Striking diversity of vmp1, a variable gene encoding a putative membrane protein of the stolbur phytoplasma. Appl Environ Microbiol 2009;75:2951-2957.

Cirillo VP, Razin S: Distribution of a phosphoenolypyruvate-dependent sugar phosphotransferase system in mycoplasms. J Bacteriol 1973;113:212-217.

Contaldo N, Bertaccini A, Paltrinieri S, Windsor HM, Windsor GD: Axenic culture of plant pathogenic phytoplasmas. Phytopathol Mediterr 2012;51:607-617.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS: Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 2002;99:5261-5266.

Dean FB, Nelson JR, Giesler TL, Lasken RS: Rapid amplification of plasmid and phage DNA using phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res 2001;11:1095-1099.

Dellaporta SL, Wood J, Hicks JB: A plant DNA minipreparation: Version ii. . Molecular Biology Reports 1983;1:3.

Deng S, Hiruki C: Amplification of 16s rRNA genes from culturable and unculturable mollicutes. Journal of Microbiological Methods 1991;14:8.

Esteban JA, Salas M, Blanco L: Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J Biol Chem 1993;268:2719-2726.

Fabre A, Danet JL, Foissac X: The stolbur phytoplasma antigenic membrane protein gene stamp is submitted to diversifying positive selection. Gene 2011;472:37-41.

Garnier M: Le phytoplasme du stolbur : Un agent ubiquiste. Comptes Rendus de l'Académie d'Agriculture de France 2000;86:7.

Gasparich GE: Spiroplasmas and phytoplasmas: Microbes associated with plant hosts. Biologicals 2010;38:193-203.

Gaurivaud P, Laigret F, Garnier M, Bové JM: Fructose utilization and pathogenicity of spiroplasma citri: Characterization of the fructose operon. Gene 2000;252:61-69.

618 Gundersen DE, Lee I-M, Rehner SA, Davis RE, Kingsbury DT: Phylogeny of mycoplasma-
619 like organisms (phytoplasmas): A basis for their classification. J Bacteriol
620 1994;176:5244-5254.
621 Hoshi A, Oshima K, Kakizawa S, Ishii Y, Ozeki J, Hashimoto M, Komatsu K, Kagiwada S,
622 Yamaji Y, Namba S: A unique virulence factor for proliferation and dwarfism in
623 plants identified from a phytopathogenic bacterium. Proc Natl Acad Sci U S A
624 2009;106:6416-6421.
625 Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken
626 RS: Unbiased whole-genome amplification directly from clinical samples. Genome
627 Res 2003;13:954-964.
628 Huang W, Li L, Myers JR, Marth GT: Art: A next-generation sequencing read simulator.
629 Bioinformatics 2012;28:593-594.
630 Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC: Integrative analysis of
631 environmental sequences using megan4. Genome Res 2011;21:1552-1560.
632 IRPCM: '*Candidatus* phytoplasma', a taxon for the wall-less, non-helical prokaryotes that
633 colonize plant phloem and insects. Int J Syst Evol Microbiol 2004;54:1243-1255.
634 James AP, Geijskes RJ, Dale JL, Harding RM: Development of a novel rolling-circle
635 amplification technique to detect banana streak virus that also discriminates
636 between integrated and episomal virus sequences. Plant Dis 2011;95:57-62.
637 Jiang YP, Lei JD, Chen TA: Purification of aster yellows agent from diseased lettuce using
638 affinity chromatography. Phytopathology 1988;78:4.
639 Kollar A, Seemüller E, Bonnet F, Saillard C, Bové JM: Isolation of the DNA of various plant
640 pathogenic mycoplasma-like organisms from infected plants. Phytopathology
641 1990;80:5.
642 Kube M, Mitrovic J, Duduk B, Rabus R, Seemüller E: Current view on phytoplasma
643 genomes and encoded metabolism. ScientificWorldJournal 2012;2012:185942.
644 Kube M, Schneider B, Kuhl H, Dandekar T, Heitmann K, Migdoll AM, Reinhardt R,
645 Seemüller E: The linear chromosome of the plant-pathogenic mycoplasma
646 '*Candidatus* Phytoplasma mali'. BMC Genomics 2008;9:306.
647 Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC: Mutation detection and
648 single-molecule counting using isothermal rolling-circle amplification. Nat Genet
649 1998;19:225-232.
650 McCoy RE, Caudwell A, Chang CJ, Chen TA, Chiykowski LN, Cousin MT, J.L. D, de Leeuw
651 GTN, Golino DA, Hackett KJ, Kirkpatrick BC, Marwitz R, Petzold H, Sinha RC,
652 Sugiura M, Whitcomb RF, Yang IL, Zhu BM, Seemüller E: Plant diseases associated
653 with mycoplasma-like organisms.; in Whitcomb RFT, J.G. (ed): The Mycoplasmas.
654 San Diego, Academic Press, 1989, vol V, pp 545-640.
655 Neriya Y, Sugawara K, Maejima K, Hashimoto M, Komatsu K, Minato N, Miura C,
656 Kakizawa S, Yamaji Y, Oshima K, Namba S: Cloning, expression analysis, and
657 sequence diversity of genes encoding two different immunodominant membrane
658 proteins in poinsettia branch-inducing phytoplasma (poibi). FEMS Microbiol Lett
659 2011;324:38-47.
660 Oshima K, Kakizawa S, Nishigawa H, Jung HY, Wei W, Suzuki S, Arashida R, Nakata D,
661 Miyata S, Ugaki M, Namba S: Reductive evolution suggested from the complete
662 genome sequence of a plant-pathogenic phytoplasma. Nat Genet 2004;36:27-29.
663 Quaglino F, Zhao Y, Casati P, Bulgari D, Bianco PA, Wei W, Davis RE: '*Candidatus*
664 Phytoplasma solani', a novel taxon associated with stolbur and bois noir related
665 diseases of plants. Int J Syst Evol Microbiol 2013.

Rabus R, Kube M, Beck A, Widdel F, Reinhardt R: Genes involved in the anaerobic degradation of ethylbenzene in a denitrifying bacterium, strain EbN1. Arch Microbiol 2002;178:506-516.
Reid SJ, Abratt VR: Sucrose utilisation in bacteria: Genetic organisation and regulation. Appl Microbiol Biotechnol 2005;67:312-321.
Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: Artemis: Sequence visualization and annotation. Bioinformatics 2000;16:944-945.
Saccardo F, Martini M, Palmano S, Ermacora P, Scortichini M, Loi N, Firrao G: Genome drafts of four phytoplasma strains of the ribosomal group 16S rIII. Microbiology 2012;158:2805-2814.
Seddas AM, R.; Daire, X.; Boudon-Padieu, E.; Caudwell, A. : Purification of grapevine flavescence dorée mlo (mycoplasma-like organism) by immunoaffinity. Current Microbiology 1993;27:8.
Seemüller E, Kampmann M, Kiss E, Schneider B: *Hflb* gene-based phytopathogenic classification of '*Candidatus* Phytoplasma mali' strains and evidence that strain composition determines virulence in multiply infected apple trees. Mol Plant Microbe Interact 2011;24:1258-1266.
Smart CD, Schneider B, Blomquist CL, Guerra LJ, Harrison NA, Ahrens U, Lorenz KH, Seemüller E, Kirkpatrick BC: Phytoplasma-specific pcr primers based on sequences of the 16S-23S rRNA spacer region. Appl Environ Microbiol 1996;62:2988-2993.
Sonnhammer EL, Durbin R: A workbench for large-scale sequence homology analysis. Comput Appl Biosci 1994;10:301-307.
Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K: Whole-genome multiple displacement amplification from single cells. Nat Protoc 2006;1:1965-1970.
Sugio A, Kingdom HN, MacLean AM, Grieve VM, Hogenhout SA: Phytoplasma protein effector Sap11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. Proc Natl Acad Sci U S A 2011;108:E1254-1263.
Tran-Nguyen LT, Gibb KS: Optimizing phytoplasma DNA purification for genome analysis. J Biomol Tech 2007;18:104-112.
Tran-Nguyen LT, Kube M, Schneider B, Reinhardt R, Gibb KS: Comparative genome analysis of '*Candidatus* Phytoplasma australiense' (subgroup tuf-australia i; rp-a) and '*Ca*. Phytoplasma asteris' strains OY-M and AY-WB. J Bacteriol 2008;190:3979-3991.
van den Broek LA, van Boxtel EL, Kievit RP, Verhoef R, Beldman G, Voragen AG: Physico-chemical and transglucosylation properties of recombinant sucrose phosphorylase from *Bifidobacterium adolescentis* DSM20083. Appl Microbiol Biotechnol 2004;65:219-227.
Vanhelden M, Tjallingii WF, Teris A, Vanbeek TA: Phloem sap collection from lettuce (*Lactuca sativa* l) - chemical comparison among collection methods. J Chem Ecol 1994;20:3191-3206.
Wei W, Davis RE, Jomantiene R, Zhao Y: Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. Proc Natl Acad Sci U S A 2008;105:11827-11832.
Xu Y, Gao S, Bruno JF, Luft BJ, Dunn JJ: Rapid detection and identification of a pathogen's DNA using phi29 DNA polymerase. Biochem Biophys Res Commun 2008;375:522-525.

715 Yilmaz S, Allgaier M, Hugenholtz P: Multiple displacement amplification compromises
716 quantitative analysis of metagenomes. Nat Methods 2010;7:943-944.
717
718