



Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma

Kenro Oshima¹, Shigeyuki Kakizawa¹, Hisashi Nishigawa¹, Hee-Young Jung¹, Wei Wei², Shiho Suzuki¹, Ryo Arashida¹, Daisuke Nakata¹, Shin-ichi Miyata¹, Masashi Ugaki^{1,2} & Shigetou Namba^{1,2}

The minimal gene set essential for life has long been sought. We report the 860-kb genome of the obligate intracellular plant pathogen phytoplasma (*Candidatus Phytoplasma asteris*, OY strain). The phytoplasma genome encodes even fewer metabolic functions than do mycoplasma genomes. It lacks the pentose phosphate cycle and, more unexpectedly, ATP-synthase subunits, which are thought to be essential for life. This may be the result of reductive evolution as a consequence of life as an intracellular parasite in a nutrient-rich environment.

The genus *Phytoplasma* (phylum Firmicutes, class Mollicutes), previously called mycoplasma-like organisms, is an economically important group of bacterial phytopathogens transmitted by insects^{1,2}. Like other mollicutes, phytoplasmas lack cell walls, are pleomorphic and reside endocellularly in insects and plant phloem. Unlike the well characterized mycoplasmas (mollicutes that reside extracellularly in their animal hosts), the phytoplasmas are endocellular. Their metabolic pathways and host interactions are of interest in agricultural and basic sciences. But the inability to culture phytoplasmas *in vitro* has hindered their characterization at the molecular level, the determination of factors involved in their pathogenicity and the discovery of effective ways to control or cure phytoplasma diseases.

The genomes of four plant-pathogenic bacteria have been completely sequenced³. All of these bacteria inhabit nutrient-poor xylem or nonspecific tissues. Here, we report the first completely sequenced genome of a phloem-limited plant-pathogenic bacterium, phytoplasma OY strain⁴ (*Candidatus Phytoplasma asteris*), line OY-M^{5,6}. Phytoplasma is the first obligate intracellular bacterium inhabiting both plant and insect to have its genome sequenced.

The genome consists of one chromosome of 860,631 bp and two smaller extrachromosomes, EcOYM and pOYM⁷ (Table 1 and Fig. 1a; Supplementary Methods, Supplementary Fig. 1 and Supplementary Table 1 online). The chromosome is a circular DNA molecule with 28% G + C content. It contains 754 open reading frames (ORFs), comprising 73% of the chromosome; 66% of these ORFs have significant homology to sequences deposited in GenBank.

Like the mycoplasma chromosome, the phytoplasma chromosome lacks many genes related to amino acid and fatty acid biosynthesis, the tricarboxylic acid cycle and oxidative phosphorylation. Unlike the mycoplasma chromosome, it also lacks genes for the phosphotransferase system and for metabolizing UDP-galactose to glucose 1-phosphate (Fig. 1b; Supplementary Tables 2 and 3 online), suggesting that phytoplasmas possess a unique sugar intake and metabolic system.

The pentose phosphate pathway (*P*) is conserved in bacteria, such as *Escherichia coli* and *Bacillus subtilis*, and in reduced-genome bacteria, such as *Buchnera* sp., an obligate symbiont of aphids, and mollicute bacteria^{8–10}. But no genes involved in this pathway are present in the phytoplasma genome. Cells use the *P* to synthesize NADPH and to obtain reducing power. The *P* also supplies the ribose 5-phosphate necessary to synthesize nucleotides. As phytoplasmas lack most of the genes necessary to synthesize nucleotides, except for those in the salvage pathway, they probably do not synthesize nucleotides, but rather absorb them from their environment, unlike mycoplasmas. As phytoplasmas inhabit the nutrient-rich host cytoplasm, they probably evolved retrogressively, losing more genes of this type than did mycoplasmas; they probably assimilate most of the metabolites that they cannot synthesize from the host cell. By contrast, the phytoplasma genome encodes folate synthesis genes, which may allow phytoplasmas to adapt to the very different plant and insect environments that they inhabit.

The phylogenetically distant *Rickettsia prowazekii* has incomplete nucleotide synthesis pathways¹¹, and phytoplasmas have a similar gene set for nucleotide metabolism (see Supplementary Table 4 online). As both phytoplasmas and *R. prowazekii* have an endocellular lifestyle, this common gene loss may reflect convergent evolution. No transport

Table 1 General features of the phytoplasma genome

	Chromosome	EcOYM	pOYM
Length (bp)	860,631	5,025	3,932
G + C content (percent)	28	25	24
Protein-coding region (percentage of chromosome)	73	71	75
Protein-coding genes			
With assigned function	446	2	2
Conserved hypothetical	51		
Hypothetical	257	4	3
Total	754	6	5
Average length of protein-coding genes (bp)	785	597	588
tRNA	32		
rRNA operons	2		

¹Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, 202 Bioscience Building, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan. ²Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. Correspondence should be addressed to S.N. (snamba@ims.u-tokyo.ac.jp).

systems for nucleosides, nucleotides or NADP/NADPH were identified in the phytoplasma genome, although they are probably present.

Most bacteria, including reduced-genome bacteria⁸⁻¹⁰, produce F₀F₁-type ATP synthase and synthesize ATP through a reverse proton discharge reaction. The F₀F₁-type ATP-synthase subunits necessary for ATP-proton motive force interconversion are essential in *Mycoplasma genitalium*^{8,12}, and all ATP-synthase subunits are members of the proposed minimal gene set indispensable for life¹³. *Chlamydia trachomatis* lacks these genes but contains the archaeal A₀A₁-type ATP synthase¹⁴. Unexpectedly, none of the eight ATP-synthase subunits was identified in the phytoplasma genome. Phytoplasma is the first living organism found to lack them all¹³. *C. trachomatis* has an ATP/ADP translocase and imports host ATP¹⁴, but the phytoplasma genome does not encode similar genes, suggesting that it imports host ATP by an unknown mechanism. Alternatively,

ATP synthesis in phytoplasma might be strongly dependent on the glycolysis pathway, as glycolysis turnover is increased in *B. subtilis* strains in which the *atp* operon has been deleted¹⁵.

The phytoplasma chromosome includes a gene encoding sucrose phosphorylase, which cleaves sucrose into glucose and fructose. But evidence that its ORF is incomplete and the absence of sucrose phosphorylase activity suggest that this gene is non-functioning; an ancestral phytoplasma may have had a functional copy of this gene (data not shown). Sucrose concentration is much lower in insect cells, which might explain why phytoplasmas no longer require this gene.

Analysis of the phytoplasma genome identified glucanase and hemolysin-like proteins as possible virulence factors. No other known virulence genes were found in the phytoplasma genome, suggesting that new mechanisms govern their virulence. Although

metabolic genes are scarce, the phytoplasma genome contains as many as 27 genes encoding transporter systems, such as malate, metal-ion and amino-acid transporters, some of which have multiple copies (Fig. 1b; Supplementary Table 5 online), suggesting that phytoplasmas aggressively import many metabolites from the host cell. The consumption of these metabolites with phytoplasma infection may greatly disturb the metabolic balance of the host cell, causing disease symptoms.

In addition to many transporter genes, there are multiple redundant copies of the genes *uvrD*, *hflB*, *tnk*, *dam* and *ssb* (18% of the total genes), which are present as single genes in many bacteria (Fig. 1a). This might explain why the phytoplasma genome is larger than that of *M. genitalium* (which is ~580 kb) despite the loss of many metabolic genes.

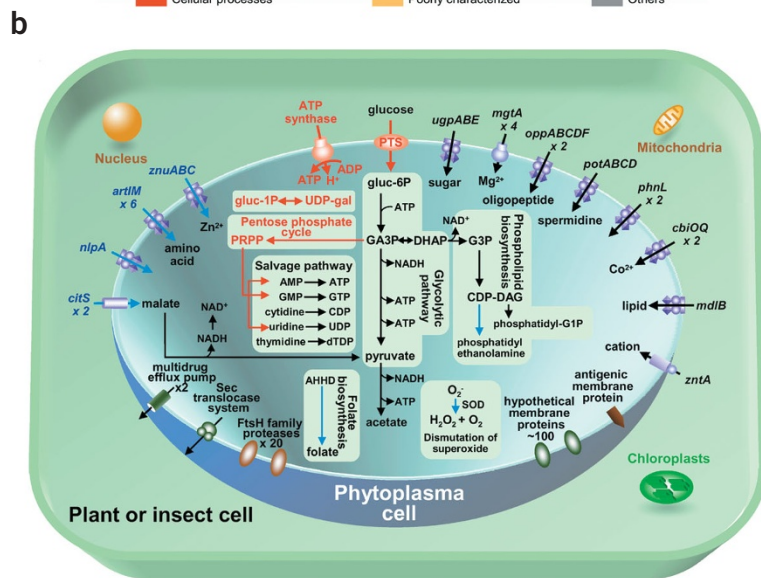
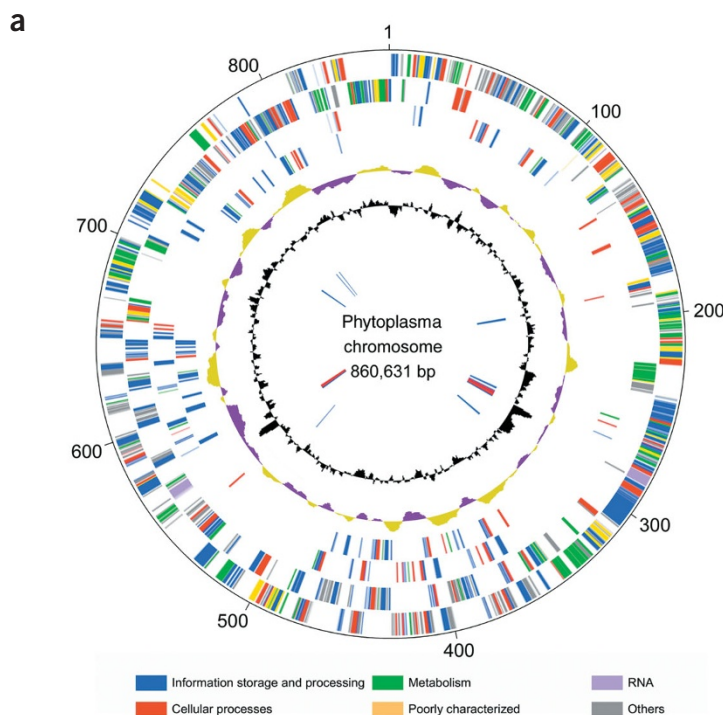


Figure 1 Schematic representation of the chromosome and metabolic pathways of phytoplasma. (a) Circular representation of the phytoplasma chromosome. The outer circle shows predicted protein-coding regions on the plus strand classified by function using the color code at the bottom of the figure. The second circle shows predicted protein-coding regions on the minus strand. The third and fourth circles show the multiple redundant genes on the plus and minus strands, respectively (see Supplementary Note online). The fifth circle shows the GC-skew value ((G - C)/(G + C)). The sixth circle shows the G + C content (higher values outward). The seventh and eighth circles show tRNAs (blue) and rRNAs (red) on the plus and minus strands, respectively. The scale in kb is indicated (numbers on the outside of the chromosome). (b) An overview of phytoplasma metabolism and transport. Shown are transporters and the main elements of metabolic pathways, deduced from the set of genes with predicted functions. Red arrows: pathways or transporters found in the *M. genitalium* genome but not in the phytoplasma genome. Blue arrows: pathways or transporters found in the phytoplasma genome but not in the *M. genitalium* genome. Detailed information about abbreviated names is given in Supplementary Note online.

In conclusion, analysis of the phytoplasma genome suggests reductive evolution as a consequence of its life as an intracellular parasite in a nutrient-rich environment and shows that phytoplasma lacks several genes previously considered to be essential for autonomously replicating cells. Instead, it possesses transporter genes involved in the uptake of substances from the cytoplasm of surrounding host cells. Although it is difficult to define the minimal genome concept for pathogens and symbionts that rely on their hosts, phytoplasmas are unique bacteria that contain a new minimal gene set.

Accession numbers. Sequences are available in the DNA Data Bank of Japan with the following accession numbers: phytoplasma chromosome, AP006628; EcOYM, AB076263; pOYM, AB061723.

URL. Detailed information about the phytoplasma genome sequencing project is available at <http://user.ecc.u-tokyo.ac.jp/~knamba/phytoplasma/>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank T. Shiomi, M. Tanaka, T. Usugi, M. Sato, H. Kawakita, N. Nishimura, T. Tsuchizaki and S. Nakajima for discussion. This work was supported partly by Grants-in-Aid of Scientific Research from the Japan Society for the Promotion of Science and a Grant-in-Aid of Scientific Research on Priority Areas (C) "Genome

Biology" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 26 September; accepted 13 November 2003

Published online at <http://www.nature.com/naturegenetics/>

1. Lee, I.M., Davis, R.E. & Gundersen-Rindal, D.E. *Annu. Rev. Microbiol.* **54**, 221–255 (2000).
2. Doi, Y., Teranaka, M., Yora, K. & Asuyama, H. *Ann. Phytopathol. Soc. Jpn.* **33**, 259–266 (1967).
3. Van Sluys, M.A. *et al. Annu. Rev. Phytopathol.* **40**, 169–189 (2002).
4. Namba, S., Oyaizu, H., Kato, S., Iwanami, S. & Tsuchizaki, T. *Int. J. Syst. Bacteriol.* **43**, 461–467 (1993).
5. Oshima, K. *et al. Phytopathology* **91**, 1024–1029 (2001).
6. Oshima, K. *et al. J. Gen. Plant Pathol.* **68**, 225–236 (2002).
7. Nishigawa, H. *et al. Microbiology* **148**, 1389–1396 (2002).
8. Fraser, C.M. *et al. Science* **270**, 397–403 (1995).
9. Razin, S., Yagev, D. & Naot, Y. *Microbiol. Mol. Biol. Rev.* **62**, 1094–1156 (1998).
10. Wernegreen, J.J. *Nat. Rev. Genet.* **3**, 850–861 (2002).
11. Andersson, S.G. *et al. Nature* **396**, 133–140 (1998).
12. Hutchison, C.A. *et al. Science* **286**, 2165–2169 (1999).
13. Mushegian, A.R. & Koonin, E.V. *Proc. Natl. Acad. Sci. USA* **93**, 10268–10273 (1996).
14. Stephens, R.S. *et al. Science* **282**, 754–759 (1998).
15. Santana, M. *et al. J. Bacteriol.* **176**: 6802–6811 (1994).