Molecular Phylogenetics and Evolution 149 (2020) 106826

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# A timetree for phytoplasmas (Mollicutes) with new insights on patterns of evolution and diversification

Yanghui Cao, Valeria Trivellone*, Christopher H. Dietrich

*Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820, USA*

ARTICLE INFO

*Keywords:*
*16S rRNA*
Bacteria
Divergence time
Hemiptera
*Map* gene
Plant pathogen

ABSTRACT

The first comprehensive timetree is presented for phytoplasmas, a diverse group of obligate intracellular bacteria restricted to phloem sieve elements of vascular plants and tissues of their hemipteran insect vectors. Maximum likelihood-based phylogenetic analysis of DNA sequence data from the *16S rRNA* and methionine aminopeptidase (*map*) genes yielded well resolved estimates of phylogenetic relationships among major phytoplasma lineages, 16Sr groups and known strains of phytoplasmas. Age estimates for divergences among two major lineages of *Mollicutes* based on a previous comprehensive bacterial timetree were used to calibrate an initial *16S* timetree. A separate timetree was estimated based on the more rapidly-evolving *map* gene, with an internal calibration based on a recent divergence within two related 16Sr phytoplasma subgroups in group 16SrV thought to have been driven by the introduction of the North American leafhopper vector *Scaphoideus titanus* Ball into Europe during the early part of the 20th century. Combining the resulting divergence time estimates into a final *16S* timetree suggests that evolutionary rates have remained relatively constant overall through the evolution of phytoplasmas and that the origin of this lineage, at ~641 million years ago (Ma), preceded the origin of land plants and hemipteran insects. Nevertheless, the crown group of phytoplasmas is estimated to have begun diversifying ~316 Ma, roughly coinciding with the origin of seed plants and Hemiptera. Some phytoplasma groups apparently associated with particular plant families or insect vector lineages generally arose more recently than their respective hosts and vectors, suggesting that vector-mediated host shifts have been an important mechanism in the evolutionary diversification of phytoplasmas. Further progress in understanding macroevolutionary patterns in phytoplasmas is hindered by large gaps in knowledge of the identity of competent vectors and lack of data on phytoplasma associations with non-economically important plants.

## 1. Introduction

Phylum *Tenericutes* includes some of the most economically important bacteria species because of their agricultural and medical importance (Bertaccini et al., 2014; Fletcher et al., 2006; Jensen, 2017; Waites, 2016). *Tenericutes* includes the class *Mollicutes*, three taxa in provisional status "*Candidatus* (*Ca.*) Izimaplasma" (Skennerton et al., 2016), and several taxa of unspecified rank. Although some authors strongly argued that *Mollicutes* should be moved to phylum *Firmicutes* (Davis et al., 2013; Ogawa et al., 2011; Yutin and Galperin, 2013), the taxonomic status of stand-alone phylum *Tenericutes* bestowed in 1984 (Murray, 1984) is strongly supported by two distinctive features that, taken together, set *Tenericutes* apart from the *Firmicutes*: the inability to synthetize precursors of peptidoglycan and, therefore, a cell wall (Brown, 2010; Skennerton et al., 2016), and extreme reduction of the genome (530–2220 kbp, Razin, 2006). *Mollicutes* includes five orders: *Mycoplasmatales*, *Entomoplasmatales*, *Haloplasmatales*, *Acholeplasmatales*, and *Anaeroplasmatales* (Parte, 2018) and some groups of taxa that still remain unclassified or poorly characterized ("*Ca.*" species, environmental samples and undescribed strains). A new order *Mycoplasmoidales* was recently proposed to include mycoplasmas in the "Hominis" and "Pneumoniae" groups, previously included in *Mycoplasmatales* (Gupta et al., 2018). Phytoplasmas, one group of bacteria considered to be *bona fide Mollicutes*, have been listed as *incertae sedis* within the order *Acholeplasmatales* (Brown et al., 2010). They include 44 '*Candidatus* Phytoplasma' species (Harrison et al., 2015; IRPCM, 2004), over 30 16S ribosomal RNA groups (Bertaccini et al., 2014; Wei et al., 2007), and 2502 environmental samples deposited in the National Center for Biotechnology Information (NCBI) database (Federhen, 2012). Some authors tentatively allocated phytoplasmas to family *Acholeplasmataceae* because, according to previous molecular phylogenies, they form a distinct clade derived from within the genus *Acholeplasma* (Martini

* Corresponding author at: 1816 South Oak Street, Champaign, IL 61820, USA.
*E-mail address:* valeria.trivellone@gmail.com (V. Trivellone).

https://doi.org/10.1016/j.ympev.2020.106826
Received 26 November 2019; Received in revised form 12 February 2020; Accepted 7 April 2020
Available online 10 April 2020

et al., 2014). Phytoplasmas were first observed using microscopy in 1967 (Doi et al., 1967) but, because they are very difficult to culture *in vitro*, their great diversity only began to be revealed in the 1990s through development of molecular techniques for characterizing groups and subgroups (Namba et al., 1993). This classification method was based on phylogenetic analysis of *16S rRNA*, with other more variable genetic loci used for finer-scale characterization. Thus, phytoplasmas are currently organized into phylogenetic groups, each designated by a Roman numeral, and different strains within a group are given a separate letter designation defining the subgroup (currently more than 120) (Lee et al., 2010; Martini et al., 2007; Pérez-López et al., 2016; Zhao et al., 2009).

Previous research on *Tenericutes* phylogeny has mostly relied on the sequences of the *16S rRNA* gene. These indicate that *Mollicutes* are a well-supported monophyletic entity (Ludwig and Klenk, 2001) having diverged from a Gram-positive ancestor by regressive evolution (Razin et al., 1998). Early *16S* phylogenies characterized five phylogenetic groups and 16 phylogenetic clusters (Pettersson et al., 1996; Weisburg et al., 1989). These are partly congruent with a more recent phylogeny based on phosphoglycerate kinase (Pgk) amino acid sequences (Wolf et al., 2004). A recent multilocus phylogeny confirmed four well-established major groups of *Tenericutes* ("*Acholeplasma*", "*Spiroplasma*", "Pneumoniae" and "Hominis") (Gupta et al., 2018), but phytoplasmas were underrepresented in this dataset. In 2007, the first extensive phylogenetic investigation to include a large sample of phytoplasma groups, other *Mollicutes* and Gram-positive bacteria showed that *Mollicutes* includes two main subclades: one comprising 46 phytoplasma strains and two *Acholeplasma* species, and the second including 14 representative *Mollicutes* from the other orders (Martini et al., 2007). The most recent phylogeny of phytoplasmas, based on *16S* and including 145 taxa and two species of *Acholeplasma*, provided further insights, suggesting that each phylogenetically distinct subgroup is equivalent to a putative species in the provisional "*Ca.* Phytoplasma" genus (Harrison et al., 2015).

*Tenericutes* bacteria have evolved a broad range of lifestyles, including free-living, commensalism and parasitism (Razin, 2006; Rivera-Tapia et al., 2002; Tully, 1996). Most of the well-known species are associated with humans or other vertebrates and are commensals inhabiting different organs. Some are suspected to be associated with diseases, whereas others are proven pathogens causing important diseases to animals (mycoplasmas) or plants (phytoplasmas and three species of *Spiroplasma*). Some free-living species are associated with inert substrates (e.g. "*Ca.* Izimaplasma") or animal/plant surfaces (e.g. *Acholeplasma laidlawii*). These observed lifestyles and host associations occur in multiple lineages of *Mollicutes* suggesting a high level of ecological plasticity in the evolution of the group as a whole. In contrast, the large phytoplasma clade represents a unique and highly specialized group defined by its adaptation to obligate intracellular parasitism. All known phytoplasmas live strictly within insects and plant phloem, have evolved mechanisms for evading their host's defenses (Tomkins et al., 2018) and use various strategies to successfully exploit the host's metabolic pathways (Chang et al., 2018; MacLean et al., 2011; Maejima et al., 2014). Most known phytoplasmas are pathogenic to plants and have been proven to be associated with devastating diseases that cause significant losses in agriculture (Bertaccini et al., 2014). In contrast, phytoplasmas evolved a commensal relationship with hemipteran (sap-sucking) insect vectors (mainly Auchenorrhyncha) that move these bacteria from plant to plant. They rarely negatively affect the fitness of insect hosts and in some cases have been shown to confer positive fitness benefits on vectors (Hogenhout et al., 2008). The highly specialized nature of phytoplasma-plant-insect associations, the widespread geographic occurrence of phytoplasmas, and the extensive phylogenetic diversity of the phytoplasma lineage suggest that this group of bacteria has been co-evolving with its plant and insect hosts for a very long time. Unfortunately, no attempts have yet been made to estimate the age of the group and its constituent lineages using molecular divergence time methods.

Because soft-bodied organisms such as bacteria are generally not preserved in the fossil record, indirect methods must be used to estimate the times of origin of such groups. Divergence times of bacterial lineages have been estimated using three indirect methods: ecological events, inference from host fossils, and inference from Eukaryotic molecular clocks (Ochman et al., 1999). An earlier evolutionary reconstruction using *16S rRNA* phylogenetic trees calibrated by correlating the nodes with ecological and geological events, showed that the lineage of *Mollicutes* from which phytoplasmas are derived diverged from ancestral *Firmicutes* about 605 million years ago (Ma) and, subsequently (407 Ma) the ancestral *Acholeplasma*-like Mollicute gave rise to two major branches, AAP (*Acholeplasma*, *Anaeroplasma*, *Asteroleplasma* and phytoplasmas) and SEM (*Spiroplasma*, *Entomoplasma*, *Mesoplasma*, *Mycoplasma* and *Ureaplasma*) (Maniloff, 2002). A comprehensive timetree using 90 calibration points from other studies and small-subunit rRNA was recently constructed for prokaryotes (Marin et al., 2016) revealing a constant overall diversification rate and suggesting that emergence of new lineages is a random process. Although non-uniform rates of molecular evolution were observed between different bacterial lineages of 42 obligate endosymbionts, the rates still appeared constant within each clade (Kuo and Ochman, 2009). Unfortunately, phytoplasmas have, so far, not been included in published bacterial timetrees.

The rapid pace of discovery of new phytoplasmas has recently yielded large amounts of relevant sequence data, providing a valuable opportunity to estimate a phytoplasma timetree and explore patterns in their diversification and phylogenetic relatedness. Previous phylogenetic analyses based on *16S rRNA* and *secY* (protein-coding gene) sequences, as well as a recent large-scale phylogenomic analysis of the bacterial phylum *Tenericutes*, strongly support the monophyly of phytoplasmas, indicating that the group was derived from within a clade comprising facultative plant- and animal-associated bacteria within class *Mollicutes* (Gupta et al., 2018; Martini et al., 2007). Thus, extant phytoplasmas apparently evolved from a single common ancestor that acquired a specific association with phloem-feeding insects and their host plants (presumably a single ancestral host plant and vector species). Several phytoplasmas are widespread and infect a wide variety of plants but many groups appear to be more restricted, with a tendency to be associated with particular biogeographic regions and/or plant families (Trivellone, 2019). The great phyletic diversity, widespread present-day geographic distribution, and association of the group with a wide variety of host species suggest that the group has been evolving in close association with its plant hosts and hemipteran vectors for millions of years.

The present study aimed to construct the first timetree for phytoplasmas using molecular divergence time methods and available DNA sequence data. The general objective was to determine whether rates of molecular divergence within the phytoplasma lineage are similar to those of related non-plant-pathogenic bacteria and to look for possible correspondence between the origins of particular phytoplasma lineages and historical biogeographic processes or events in the evolution of their plant and hemipteran insect hosts.

## 2. Materials and methods

### 2.1. Data sets

To estimate the divergence times of phytoplasmas, two genetic datasets with varying taxonomic sampling were used, one for *16S rRNA* and one for the *methionine aminopeptidase* (*map*) gene.

The *16S rRNA* dataset (*16S* hereafter) consisted of 220 sequences obtained from the NCBI database. The ingroup included 169 designated phytoplasma subgroups (44 described as *Ca.* Phytoplasma (P.) species, including incidental citations), except 16SrIII-O, 16SrIII-R, 16SrVI-G, 16SrVII-F, and 16SrX-E subgroups. The outgroup included 49 taxa

representing all the orders of *Tenericutes* (15 *Acholeplasma* species, three species of *Anaeroplasmatales*, one *Haloplasmatales*, 20 *Mycoplasmatales*, seven *Entomoplasmatales*, two "*Candidatus* Izimaplasma" species, one "*Candidatus* Hepatoplasma" species), and two species of *Firmicutes* (*Bacillus pumilus* and *Clostridioides difficile*) used to root the tree. These outgroup taxa were chosen to represent diverse geographic origins, lifestyles and association with hosts (Table S1).

The *map* dataset was assembled with 82 phytoplasma strains and the outgroup *Acholeplasma palmae*. The ingroup consists of seven strains in the phylogenetic group 16SrI (including five subgroups), four in 16SrII (three), seven in 16SrIII (six), two in 16SrIV (two), 47 in 16SrV (five), four in 16SrVI (two), two in 16SrX (two), four in 16SrXII (three), and single subgroups and strains from 16SrVII, 16SrVIII, 16SrIX, 16SrXI, 16SrXIII and 16SrXVIII groups. The 47 strains in the 16SrV include 10 "*Ca.* P. ulmi" strains (16SrV-A), two "*Ca.* P. ziziphi" (16SrV-B), two "*Ca.* P. rubi" (16SrV-E), 32 strains in the subgroups 16SrV-C and -D, and one strain of undesignated subgroup (Table S2).

### 2.2. Phylogenetic analysis

Sequence alignment was conducted using the Muscle algorithm as implemented in MEGA v7.0.26 (Kumar et al., 2016) with default settings, followed by manual editing. The *map* alignment was trimmed according to the length most commonly found for published 16SrV strains. The aligned dataset is available at https://zenodo.org under DOI https://doi.org//10.5281/zenodo.3366176. DNA evolution models and partitioning schemes were determined by PartitionFinder 2 (Lanfear et al., 2014). For both *16S rRNA* and *map*, the GTR + G + I model, which considers variable base frequencies and the proportion of invariant sites, had the lowest corrected Akaike Information Criterion score during model selection and thus was determined as the best-fit model. The maximum likelihood (ML) tree for each gene was constructed using RAxML 8.2.11 (Stamatakis, 2014) with the "-f a" algorithm, and support for nodes was evaluated using 1000 bootstrap replicates.

### 2.3. Divergence time estimation

The RelTime algorithm implemented in MEGA 7.0.26 was used to estimate the divergence times of the phytoplasma lineage (Mello, 2018). Because phytoplasmas have not been included in previous molecular divergence time analyses, it is not known whether their evolutionary rates are the same as in lineages of related bacteria. Thus, the RelTime algorithm that allows substitution rates to vary among lineages was chosen (Tamura et al., 2018). This algorithm was used recently to estimate divergence times of the major lineages of prokaryotes overall (Marin et al., 2016).

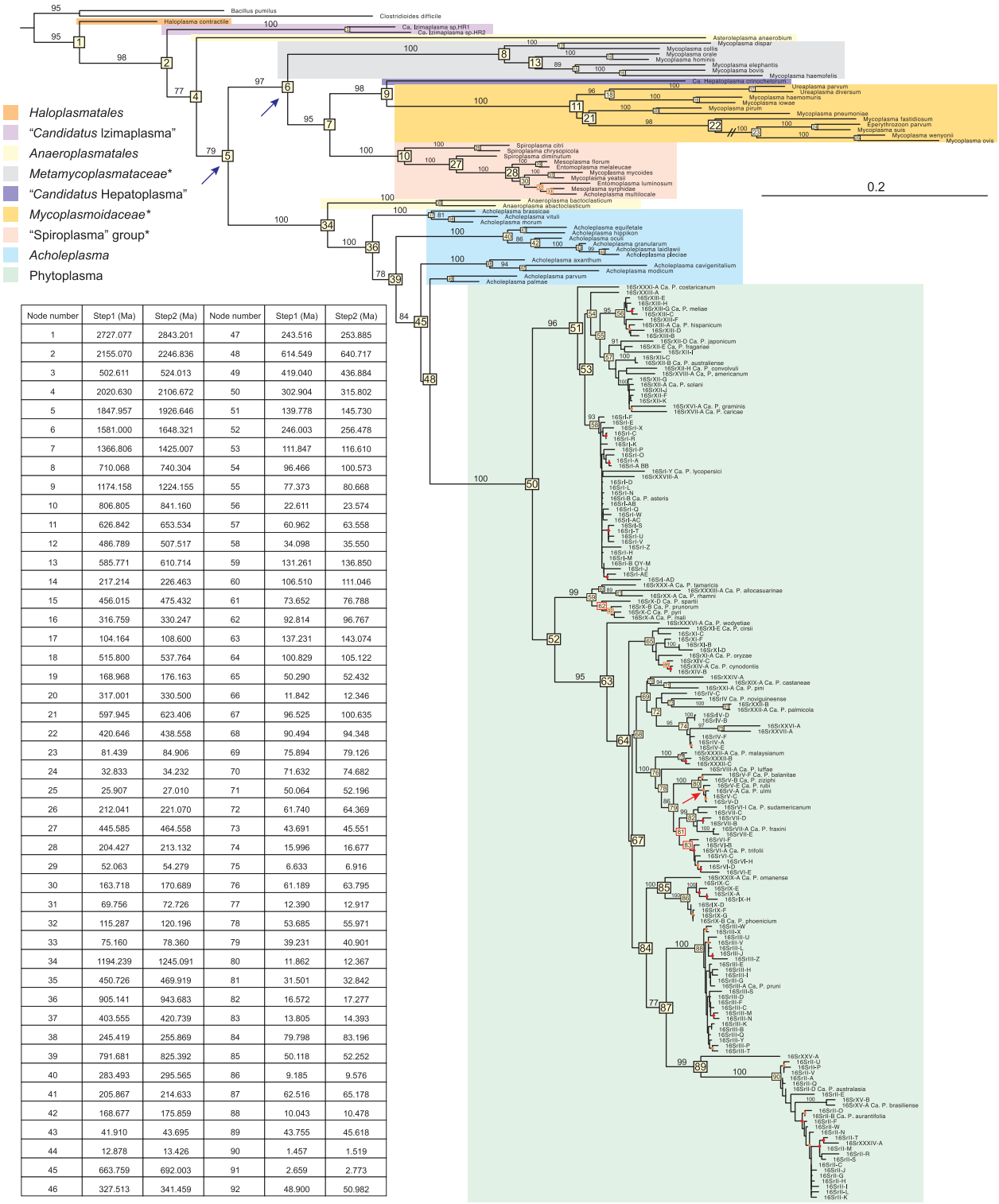
### 2.4. Calibration of the 16S timetree was carried out in two separate steps.

*Step 1*. Divergence times of phytoplasmas were initially estimated based on the *16S* tree with the ages of the two deepest nodes being constrained (Fig. 1, blue arrows) using calibration information acquired from the TimeTree web resource (http://www.timetree.org/, Kumar et al., 2017): 1) the split of *Anaeroplasmatales* + *Acholeplasmatales* from *Mycoplasmatales* + *Entomoplasmatales* was constrained at 1704.5–2070 million years ago (Ma) (Marin et al., 2016; Sheridan et al., 2003); 2) the split of *Mycoplasmatales* from *Entomoplasmatales* was constrained at 1581–1679.4 Ma (Marin et al., 2016; Sjöstrand et al., 2014).

*Step 2*. Because evolutionary rates may differ between phytoplasmas and other lineages of *Tenericutes*, an internal calibration on the *16Sr* tree based on an inferred recent divergence event in the 16SrV group was added; specifically, a cluster which includes 32 strains belonging to the 16SrV-C and -D subgroups (Fig. S1, box in grey). Some of these strains are associated with an epidemic disease named Flavescence dorée (FD) that appeared recently and threatens the wine industry of Europe. This disease and the associated phytoplasma strains have been studied intensively and, therefore, data available from the literature provide strong justification for using recent divergences among FD-related strains to calibrate the phytoplasma timetree based on ecological events. Outbreaks of FD were first observed in southwestern France in the 1950s, although FD-like symptoms had been reported as early as the 1920s (Caudwell, 1957; Caudwell, 1964). Earlier research revealed a high genetic diversity between phytoplasma strains belonging to the 16SrV-C and -D subgroups varying with host plant, strains associated with grapevine being less variable than those detected on other plants (Angelini et al., 2001, 2003; Arnaud et al., 2007). Using a multilocus sequence typing approach, it was possible to further classify the strains infecting grapevine and strictly transmitted by the introduced Nearctic leafhopper vector *Scaphoideus titanus* Ball in three consistent lineages (hereafter referred to as FD phytoplasmas *sensu stricto*; FDp according to Angelini et al., 2004): the low-variability cluster which includes the reference strain recorded in grapevine for the first time in France (FD70; Caudwell et al., 1970), the cluster including the clonal strains currently widespread in Central Europe (FD-D; Angelini et al., 2001; Daire et al., 1997), and the more-variable cluster detected mainly in Italy and in Eastern Europe (FD-C; Angelini et al., 2001). Several other European FD-related phytoplasma (FD-rp) strains infecting mainly alder, clematis and other arboreal plants including grapevine, such as those causing Alder yellows disease (e.g. ALY strain) and Palatinate grapevine yellows (PGY strains), showed a non-monophyletic origin (Arnaud et al., 2007). Fig. 2 depicts the timeline of the important anthropogenic events and scientific evidence upon which our assumption about the probable time range of FDp diversification is based. *Scaphoideus titanus* was recorded in Europe for the first time at the end of the 1950s (Bonfils and Schvester, 1960), and characterized as a highly efficient vector of FDp (Schvester et al., 1962), strictly associated with grapevine. Earlier research suggested that the European population of this leafhopper originated from a single accidental introduction (Papura et al., 2012). Introduction of *S. titanus* into Europe is thought to have occurred between 1900 and 1920, during the extensive importation of American vine rootstocks (Chuche and Thiéry, 2014) used to provide resistance to *Phylloxera* damage (Ollat et al., 2016). Phylogenetic analyses of FDp and FD-rp strains showed a high diversity of genotypes in endemic European alders and, to a lesser extent, in clematis, with some of the latter identical to strains detected in grapevines. This supports the hypothesis that alder and clematis could be the original reservoirs for recently evolved strains associated with grapevines (Angelini et al., 2004; Malembic-Maher et al., 2017). Based on *map* genetic clustering, Arnaud et al. (2007) hypothesized that spreading of FD-rp from alder (or other plant hosts) to grapevine happened at least three times, giving rise to the three FDp clusters. It seems likely that the introduction of *S. titanus* into Europe remained unnoticed until this competent vector initiated vine-to-vine transmission within vineyards, leading to evolutionary divergence of new grapevine-specific FDp strains and disease outbreaks as a result of vector-mediated isolation on the new host (grapevine).

Due to the limited availability of *16S* sequences and their high similarity among the strains of 16SrV-C and -D subgroups, it was necessary to use the more variable and rapidly evolving *map* gene (Arnaud et al. 2007) to estimate precisely the very recent divergence times between FDp and FD-rp assumed to be recently diverged phytoplasma lineages. Although other variable and rapidly evolving phytoplasma genes have been reported (e.g., *secY* gene, Lee et al., 2010; *uvrB-degV*, Malembic-Maher et al., 2011), the *map* gene was chosen because it is considered a neutral marker and sequences are available for many FDp strains.

To calibrate the *map* tree, age of the crown group of phytoplasmas inferred from *Step 1* (302.90 Ma; Fig. 1, node 50) was used. Due to the requirement of both upper and lower bounds of the calibration in MEGA, the time interval was estimated by presuming a normal distribution (although MEGA does not support use of distribution models




| Node number | Step1 (Ma) | Step2 (Ma) | Node number | Step1 (Ma) | Step2 (Ma) |
|---|---|---|---|---|---|
| 1 | 2727.077 | 2843.201 | 47 | 243.516 | 253.885 |
| 2 | 2155.070 | 2246.836 | 48 | 614.549 | 640.717 |
| 3 | 502.611 | 524.013 | 49 | 419.040 | 436.884 |
| 4 | 2020.630 | 2106.672 | 50 | 302.904 | 315.802 |
| 5 | 1847.957 | 1926.646 | 51 | 139.778 | 145.730 |
| 6 | 1581.000 | 1648.321 | 52 | 246.003 | 256.478 |
| 7 | 1366.806 | 1425.007 | 53 | 111.847 | 116.610 |
| 8 | 710.068 | 740.304 | 54 | 96.466 | 100.573 |
| 9 | 1174.158 | 1224.155 | 55 | 77.373 | 80.668 |
| 10 | 806.805 | 841.160 | 56 | 22.611 | 23.574 |
| 11 | 626.842 | 653.534 | 57 | 60.962 | 63.558 |
| 12 | 486.789 | 507.517 | 58 | 34.098 | 35.550 |
| 13 | 585.771 | 610.714 | 59 | 131.261 | 136.850 |
| 14 | 217.214 | 226.463 | 60 | 106.510 | 111.046 |
| 15 | 456.015 | 475.432 | 61 | 73.652 | 76.788 |
| 16 | 316.759 | 330.247 | 62 | 92.814 | 96.767 |
| 17 | 104.164 | 108.600 | 63 | 137.231 | 143.074 |
| 18 | 515.800 | 537.764 | 64 | 100.829 | 105.122 |
| 19 | 168.968 | 176.163 | 65 | 50.290 | 52.432 |
| 20 | 317.001 | 330.500 | 66 | 11.842 | 12.346 |
| 21 | 597.945 | 623.406 | 67 | 96.525 | 100.635 |
| 22 | 420.646 | 438.558 | 68 | 90.494 | 94.348 |
| 23 | 81.439 | 84.906 | 69 | 75.894 | 79.126 |
| 24 | 32.833 | 34.232 | 70 | 71.632 | 74.682 |
| 25 | 25.907 | 27.010 | 71 | 50.064 | 52.196 |
| 26 | 212.041 | 221.070 | 72 | 61.740 | 64.369 |
| 27 | 445.585 | 464.558 | 73 | 43.691 | 45.551 |
| 28 | 204.427 | 213.132 | 74 | 15.996 | 16.677 |
| 29 | 52.063 | 54.279 | 75 | 6.633 | 6.916 |
| 30 | 163.718 | 170.689 | 76 | 61.189 | 63.795 |
| 31 | 69.756 | 72.726 | 77 | 12.390 | 12.917 |
| 32 | 115.287 | 120.196 | 78 | 53.685 | 55.971 |
| 33 | 75.160 | 78.360 | 79 | 39.231 | 40.901 |
| 34 | 1194.239 | 1245.091 | 80 | 11.862 | 12.367 |
| 35 | 450.726 | 469.919 | 81 | 31.501 | 32.842 |
| 36 | 905.141 | 943.683 | 82 | 16.572 | 17.277 |
| 37 | 403.555 | 420.739 | 83 | 13.805 | 14.393 |
| 38 | 245.419 | 255.869 | 84 | 79.798 | 83.196 |
| 39 | 791.681 | 825.392 | 85 | 50.118 | 52.252 |
| 40 | 283.493 | 295.565 | 86 | 9.185 | 9.576 |
| 41 | 205.867 | 214.633 | 87 | 62.516 | 65.178 |
| 42 | 168.677 | 175.859 | 88 | 10.043 | 10.478 |
| 43 | 41.910 | 43.695 | 89 | 43.755 | 45.618 |
| 44 | 12.878 | 13.426 | 90 | 1.457 | 1.519 |
| 45 | 663.759 | 692.003 | 91 | 2.659 | 2.773 |
| 46 | 327.513 | 341.459 | 92 | 48.900 | 50.982 |

**Fig. 1.** Maximum likelihood tree based on *16S rRNA* gene. Bootstrap values (> 75) are shown above the branches or indicated at nodes by colored dots or boxes: red refers to 90–100 and orange refers to 75–89. The scale bar indicates the branch length, except the branch marked with "//" which was shortened, with the real length 0.4441. Blue arrows show the calibrations applied for both timetrees of *16S rRNA*, and red arrow shows the calibration used only on the final timetree of *16S*. * Families or groups *sensu* Gupta et al. (2018). Divergence times of the numbered nodes are shown in the table on the left: "Step1" column ages estimated from the preliminary *16S* timetree; "Step2" ages estimated from final *16S* timetree. Detailed information on reference strains is reported in Table S1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the calibration) with 302.90 Ma as mean and a standard deviation (SD) of 10 Ma (using the method of Sanglas et al., 2017). The 95% confidence interval 283.30–322.50 Ma was used as calibration boundaries. Then, the split between ALY + PGY-B (FD-rp strains) and the clade of FDp including nine strains in 16SrV-D subgroup and the SI04-S4 strain (16SrV-C subgroup) was constrained (Fig. S1, blue arrow) to a time range of 22–100 years before present. The minimum age (22 years) refers to the earliest published record of the strains in the

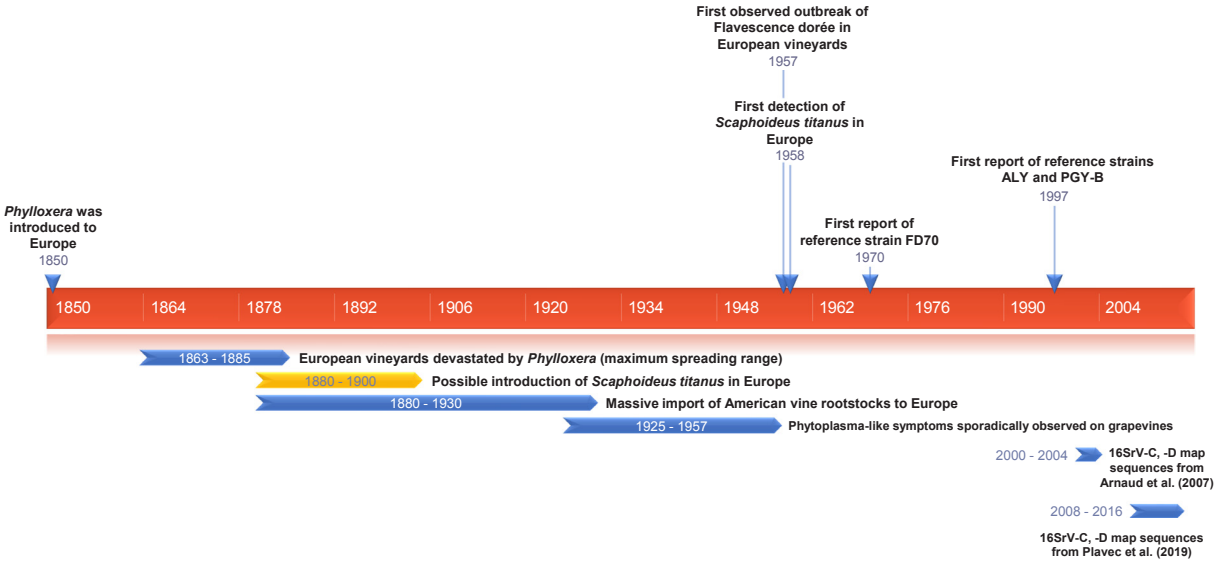



**Fig. 2.** Timeline of events supporting hypothesis of a recent diversification in phytoplasma subgroup 16SrV-C and -D corresponding to the introduction into Europe of the Nearctic vector *Scaphoideus titanus*. The records of the reference strains used in this study were reported from 1970 to 2016.

constrained clade (ALY and PGY-B), and the maximum age (100) was based on the hypothesized time of introduction of *S. titanus* into European vineyard agroecosystems (1900–1920), also corresponding with the earliest report of symptoms associated with grapevine yellows diseases observed in Europe. Although it is possible that some or all of these early disease outbreaks may have been due to a different, endemic, grapevine yellows disease (Bois noir), it seems reasonable to assume that vector-mediated divergence of the FDp phytoplasma strains began soon after introduction of the non-native leafhopper vector. Constraining the divergence of FDp strains to a narrower range with a more recent maximum age would be expected to yield slightly more recent divergence time estimates with narrower 95% confidence intervals, but we opted to apply a broader, more conservative date range calibration, given the uncertainty around the actual time of origin of the FDp strains.

To build the final *16S* timetree, the divergence time for the split of 16SrV-A, 16SrV-C, 16SrV-D and 16SrV-E estimated from the *map* tree (5.74 Ma; Fig. S1, node 23) was added as a recent internal calibration to the *16S* tree (Fig. 1, red arrow). Calibration boundaries were set as 5.16–6.33 Ma, which was the 95% confidence interval of the normal distribution with a mean of 5.74 Ma and SD of 0.3 Ma. The deeper node calibrations used in *Step 1* were also applied in this step (Fig. 1, blue arrows). Applying this additional node calibration near the tip of the phytoplasma tree enabled us to examine the effect of including an internal calibration on the divergence time estimates for phytoplasmas overall.

The final *16S* timetree was plotted against geological time using the **strap** package (Bell and Lloyd, 2015) in R 3.6.0 (R Core Team, 2019).

## 3. Results and discussion

### 3.1. Phylogenetic analysis

The *16S rRNA* gene sequence data set included 169 taxa and represented all described "*Ca.* Phytoplasma" species and almost all designated subgroups in the 35 16Sr phytoplasma groups. The *map* sequences were collected from 33 species and subgroups, representing 14 16Sr groups. Group 16SrV was sampled to include multiple strains of subgroups 16SrV-C and -D. Alignments of *16S rRNA* and *map* consisted of 1655 positions and 564 positions (both including gaps), respectively.

Relationships among *Mollicutes* recovered by the *16S* ML tree (Fig. 1) are largely congruent with previous results based on genome data (Gupta et al., 2018). "*Candidatus* Izimaplasma" (including two strains, HR1 and HR2) branches near the root and is sister to the other *Mollicutes* (excluding *Haloplasma*) although it was previously suggested as a sister clade of *Acholeplasmatales* (Skennerton et al., 2016). *Anaeroplasmatales* is polyphyletic, with *Asteroleplasma* branching more deeply than *Anaeroplasma*. Our results are in agreement with Weisburg et al. (1989), who showed that *Asteroleplasma* branched from other *Firmicutes* independently of *Anaeroplasma*. *Metamycoplasmataceae* and *Mycoplasmoidaceae sensu* Gupta et al. (2018) were recovered here as monophyletic, but contrary to Gupta et al. the latter family and "*Ca.* Hepatoplasma" was sister to the "*Spiroplasma*" group *sensu* Gupta et al. The entire clade including *Metamycoplasmataceae*, *Mycoplasmoidaceae*, "*Ca.* Hepatoplasma" and *"Spiroplasma"* group is sister to *Acholeplasmatales* + *Anaeroplasma*. Minor discrepancies among outgroup relationships compared to previous analyses are probably artifacts of low sample coverage in our dataset and were not expected to affect the divergence time estimates for phytoplasmas.

*Acholeplasmatales* was recovered as sister to *Anaeroplasma* in agreement with previous studies (Gupta et al., 2018; Johansson et al., 1998; Maniloff, 2002). Phytoplasmas are a well supported monophyletic group in *Acholeplasmatales*, forming a sister lineage to *Acholeplasma palmae* + *A. parvum*. Although branch support for this relationship is low, this result is congruent with a previous phylogeny based on genome data (Gupta et al., 2018). Interestingly, a previous phylogeny of *Acholeplasma*, not considering phytoplasmas, reported this group as monophyletic (Volokhov et al., 2007); however, when phytoplasmas are included, *Acholeplasma* is rendered paraphyletic (Gundersen et al., 1994; Gupta et al., 2018). The phytoplasma clade is separated into two major subclades, in agreement with previous results (Lee et al., 2010). The first subclade is composed of 16SrI, 16SrXII, 16SrXIII, 16SrXVI, 16SrXVII, 16SrXVIII, 16SrXXIII, 16SrXXVIII and 16SrXXXI groups with high branch support (bootstrap value 96). The remaining 16Sr subgroups form a second subclade. The analysis also strongly supports the monophyly of several 16Sr groups, including 16SrIII, 16SrV, 16SrIX, 16SrX, 16SrXIII, 16SrXIV, 16SrXV, 16SrXXII and 16SrXXXII. However, two of the largest groups 16SrI and 16SrII are polyphyletic. All the subgroups of 16SrI group together in a single clade but this group gave rise to 16SrXXVIII. The 16SrII lineage (also containing 16SrXV and 16SrXXXIV) and the 16SrXXV branch group together, forming a sister relationship to 16SrIII. Group 16SrV is sister to the lineage comprising 16SrVI and 16SrVII. Sister relationships between some other small groups are also strongly supported, including

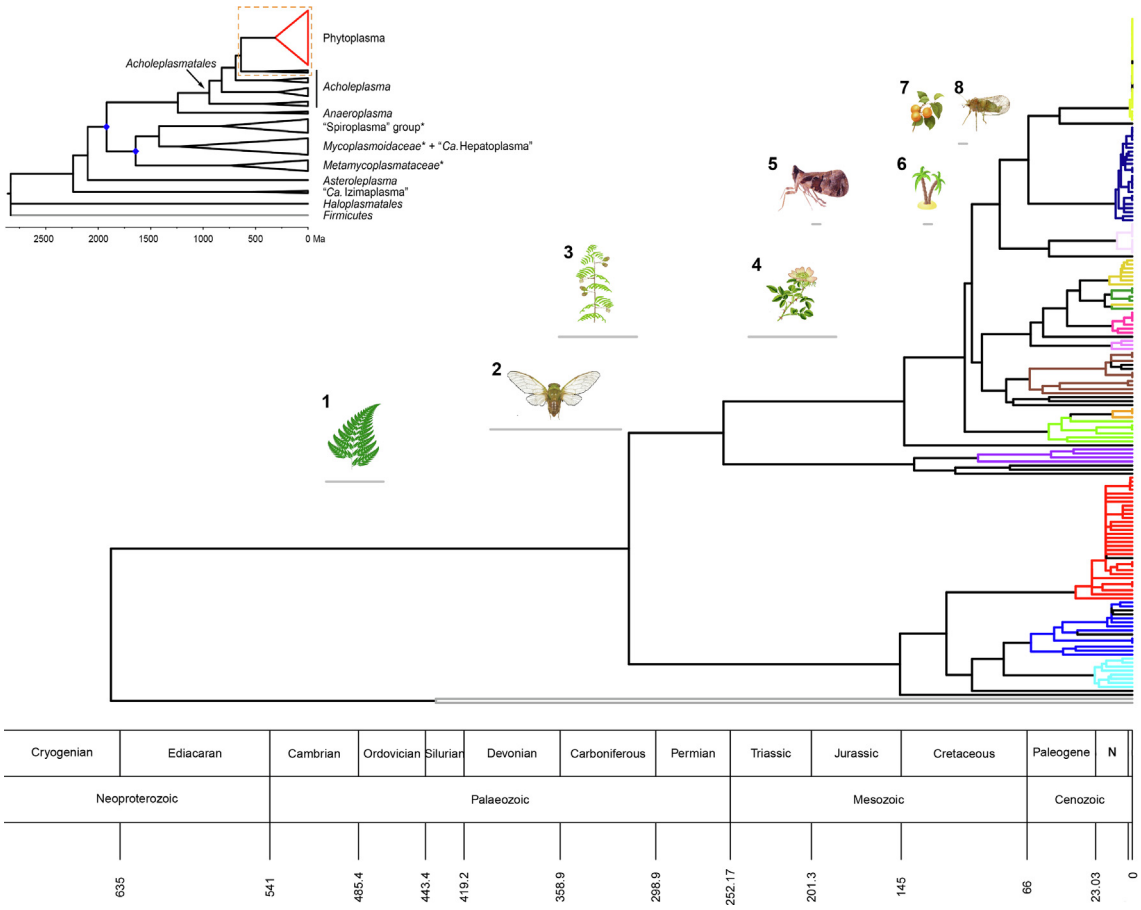



**Fig. 3.** Timetree chronogram for the final *16S rRNA* gene dataset (tree on the top left). Blue dots indicate the calibrated deeper nodes. * Families or group *sensu* Gupta et al. (2018). The divergence time estimation of phytoplasmas and *Acholeplasma* sister clade (dashed orange box) is magnified below. The horizontal scale indicates age in millions of years, with vertical lines indicating boundaries between geological time units (Era and Period). **N**: Neogene. Node age estimates correspond to those in Fig. 1, column "Step2". Branches are colored from bottom to top as follows: gray, *Acoleplasma palmae* + *A. parvum*; light blue, 16SrXIII group; blue, 16SrXII; red, 16SrI; violet, 16SrX; green, 16SrXI; orange, 16SrXIV; brown, 16SrIV + 16SrXXII; pink, 16SrXXXII; deep pink, 16SrV; dark green, 16SrVII; yellow, 16SrVI; light pink 16SrIX; navy, 16SrIII; light yellow, 16SrII. Numbered pictures: **1**, land plants (473.5–514.8 Ma; Morris et al., 2018); **2**, hemipteran insects (~300–386 Ma; Johnson et al., 2018); **3**, Spermatophyta (289–365 Ma; Kumar et al., 2017; Morris et al., 2018); **4**, angiosperms (168–246 Ma; Kumar et al., 2017; Morris et al., 2018); **5**, Fulgoroidea (~200 Ma; Johnson et al., 2018); **6**, Arecaceae (97–110 Ma; Percy et al., 2018); **7**, Rosaceae (~101.6 Ma; Xiang et al., 2017); **8**, Psyllidae ca. 95 Ma; Johnson et al., 2018; Percy et al., 2018). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

16SrIX + 16SrXXIX, 16SrXIX + 16SrXXI, 16SrXX + 16SrXXXIII and 16SrXXVI + 16SrXXVII. To our knowledge, this is the first phylogeny of phytoplasmas that includes a well sampled outgroup including representatives of all major lineages of *Tenericutes*.

The ML tree for *map* (Fig. S1) was constructed based on partial coding sequences with incomplete sampling of 16Sr phytoplasma groups and subgroups, representing only groups 16SrI to 16SrXIII and 16SrXVIII. Although the *map* sequences were not available for all subgroups considered, the *map* tree recovered two major subclades similar to the *16S* tree (Fig. 1), but with the position of 16SrX shifted from one subclade to another. Notably, neither the *16S* nor the *map* tree indicated high branch support for the position of 16SrX, but the major subclades were strongly supported when 16SrX (or the small clade containing 16SrX) was excluded. 16SrI, 16SrII, 16SrIII, 16SrIV and 16SrVI were recovered as well-supported monophyletic groups, probably due to the limited sample coverage. Group 16SrXVIII was nested within 16SrXII as in the 16S tree. Group 16SrV was also recovered as a monophyletic lineage but with moderate support (bootstrap value 72). Subgroups 16SrV-A, -C, -D and -E formed a sister group to 16SrV-B (or 16SrV-B + 16SrV-F in 16S tree), which was previously reported as the most divergent lineage with respect to the other groups (Arnaud et al., 2007; Jung et al., 2003). The positions of 16SrV-A and 16SrV-E were different in the two ML trees. 16SrV-C and -D, including FDp and FD-rp strains, were consistently recovered as closely related subgroups (clade in the gray box, Fig. S1), and three main clusters (A, B and C in Fig. S1) were detected. The first two clusters (A and B) include FDp in subgroup 16SrV-C and FD-rp strains, whereas the third one (cluster C) includes a well-supported monophyletic lineage of strains belonging to the 16SrV-D subgroup (bootstrap value 98) sister to SI04-S4 strain, and siblings to the FD-rp strains ALY + PGY-B. The ML general topology is consistent with earlier studies of the *map* gene (Arnaud et al., 2007; Plavec et al., 2019).

### 3.2. Divergence times

To overcome the absence of a fossil record for most prokaryotes, secondary calibrations and divergence times of hosts are often employed in the estimation of microbial evolutionary history (Chriki-Adeeb and Chriki, 2016; Gruen et al., 2019). Previous studies on the divergence times of *Mollicutes* (Marin et al., 2016; Sheridan et al., 2003; Sjöstrand et al., 2014) and multilocus genetic characterization of FDp and FD-rp strains supporting the assumption of recent divergences of FDp (Angelini et al., 2003; Arnaud et al., 2007; Malembic-Maher et al., 2017), provided an opportunity to illuminate the evolutionary history of phytoplasmas. In this paper, times of origin and major divergences of phytoplasmas were estimated using the *16S rRNA* gene. This molecular

marker has been widely used for phylogenetic and taxonomic classification of prokaryotes (Johansson et al., 1998).

Our initial timetree estimate, using calibrations of deeper nodes only (Fig. 1, blue arrows), suggested the divergence between the most recent common ancestor of phytoplasmas and *A. palmae* + *A. parvum* occurred ~614.55 Ma. In a previous study that attempted to infer times of origin of various bacterial lineages based on known geological events without incorporating information on molecular divergence, the split between phytoplasma and *Acholeplasma* clade was dated much later (180 Ma; Maniloff, 2002). In our tree, the crown clade of the phytoplasmas was dated ~302.90 Ma. Within the crown clade, the first major subclade (including 16SrI, 16SrXII, 16SrXIII and 6 other small groups) began to diversify in Lower Cretaceous, and the second one (including 16SrII and 16SrIII, and 24 other groups) in the Late Paleozoic (Fig. 1, column " Step 1").

In the *map* timetree, the first (including 16SrI, 16SrXII, etc.) and second (including 16SrII, 16SrIII, 16SrV, etc.) major phytoplasma subclades were dated ~ 269.71 Ma (Fig. S1, node 2) and ~174.63 Ma (Fig. S1, node 3), respectively, substantially different compared to the preliminary 16S tree: 139.77 Ma (Fig. S1, "Step1"; node 51) for the first subclade and 246 Ma (Fig. S1, "Step1"; node 52) for the second subclade. This discrepancy may be partly due to the different positions of the 16SrX group between the two trees. The age of the 16SrV lineage was estimated at ~ 15 Ma (Fig. S1, node 22) and the divergence among 16SrV-A, 16SrV-C, 16SrV-D and 16SrV-E started ~ 5.74 Ma (Fig. S1, node 23). Due to the absence of 16SrV-F in the *map* tree, the divergence time of 16SrV-A + (16SrV-E + (16SrV-C + 16SrV-D)) (Fig. S1, red arrow) rather than the crown of 16SrV was used to constrain the age of the corresponding node in the *16S* tree (Fig. 1, red arrow).

The final *16S* timetree, with the internal calibration for 16SrV applied, yielded similar divergence times to those obtained for the initial *16S* tree without the internal calibration (Fig. 1, compare columns "Step1" and "Step2"), suggesting that rates of evolution have remained relatively constant overall within this lineage.

Fig. 3 shows a simplified version of the final *16S* timetree of the phytoplasma lineage plotted against geological time with important evolutionary events indicated. The phytoplasma lineage was found to have diverged from the A. *palmae* + A. *parvum* subclade in the late Proterozoic (~641 Ma), before the emergence of land plants, dated at middle Cambrian-Early Ordovician (473.5–514.8 Ma; Morris et al., 2018), and the Hexapoda (insect) clade dated at Upper Cambrian (~500 Ma; Giribet and Edgecombe, 2019). Known *Acholeplasma* species are broadly associated with vertebrates, except for *A. pleciae* associated to *Plecia* sp. (Diptera) and *A. laidlawii* and *A. morum* associated with two species of mosquito (Brown and Johansson, 2010). So far, there is no evidence for parasitic associations of *Acholeplasma* with their hosts. We speculate that the ecological niche of the *Acholeplasma*-like ancestor of modern phytoplasmas was similar to that of extant *Acholeplasma*. Possibly these ancestral stem-group "phytoplasmas" were associated with ancestral Chordata (including vertebrates) or Panarthropoda (dos Reis et al., 2015; Edgecombe and Legg, 2014). The possibility must also be considered that the recent timetree of prokaryotes (Marin et al., 2016), upon which our calibration of the divergence of *Acholeplasmatales* from related *Mollicutes* was based, overestimated the ages of *Mollicutes* lineages included in that study. Further analysis with additional calibration points may be expected to improve the divergence time estimates presented here.

Interestingly, the basal divergence in the crown clade of phytoplasmas (315.80 Ma; Fig. 1, node 50) corresponds to recent date estimates for the rise of seed plants (Spermatophyta) 289–365 Ma (Kumar et al., 2017; Morris et al., 2018) and hemipteran insects ~ 300–386 Ma (Johnson et al., 2018). Subsequently, a few additional divergences of major phytoplasma lineages occurred between the Carboniferous and the Cretaceous but most of the recognized modern 16Sr phytoplasma groups did not appear until later in the Cretaceous or during the Cenozoic, after the radiation of angiosperms (168–246 Ma; Kumar et al., 2017; Morris et al., 2018) was underway. Many of the numbered 16S groups radiated between 20 and 50 Ma.

The highly specialized 16SrX group began a rapid radiation in the Upper Cretaceous (96.76 Ma). Within this group, subclade 16SrX-A + -B + -C (originating 50.98 Ma) is strictly associated with a group of Palearctic insect vectors in the genus *Cacopsylla* (Psyllidae) and the plant family Rosaceae. The crown of the 16SrX group was estimated ~ 96.76 Ma (Fig. 1, node 62), similarly Rosaceae emerged in Lower Cretaceous (~101.6 Ma; Xiang et al., 2017). A recent molecular phylogeny of Psylloidea recovered Psyllidae (*sensu stricto*), including *Heteropsylla* and *Cacopsylla*, in the terminal clade (Percy et al., 2018). The hemipteroid timetree of Johnson et al. (2018) did not include *Cacopsylla* but the related psyllid genus *Heteropsylla* diverged from other Psylloidea at ca. 95 Ma, slightly more recently than the estimated time of origin of Rosaceae. *Cacopsylla* comprises two well-supported subclades, the first one including the European species of *Cacopsylla* predominantly associated with Rosaceae. Interestingly, all the species of this subclade are recorded as competent or potential vectors of phytoplasmas in the Rosaceae-associated 16SrX sub-clade. The phytoplasma lineage that includes group 16SrX, the only group associated with non-Auchenorrhyncha vectors (Sternorrhyncha: Psyllidae), is one of the earliest diverging lineages among the modern phytoplasma groups, suggesting that the association of this lineage with Sternorrhyncha may be much older. Sternorrhyncha diverged from other Hemiptera ca. 386 Ma and Psylloidea diverged from other Sternorrhyncha ca. 356 Ma, predating the age (~256 Ma) of the initial divergence of the clade which includes the 16SrX group (and three small groups for which vectors are currently unknown).

The non-monophyletic group which includes 16SrIV, 16SrXXII-A, -B, 16SrXXVI-A, and 16SrXXVII-A recently evolved another specific association with plants (64.37 Ma; Fig. 1, node 72). All the strains, except 16SrXXVI-A and 16SrXXVII-A, are associated with the palm family (Arecaceae), and the known geographic distribution of this group of phytoplasmas roughly coincides with the pantropical distribution of palms. The split of Arecaceae from their closest relative is dated at 120 Ma and the divergence of the extant lineages of these plants began between 110 and 97 Ma (Anderson and Janssen, 2009). The known vectors of the two subclades in phytoplasma group IV belong to Fulgoroidea (planthoppers) which emerged ~ 200 Ma (Johnson et al., 2018), well before the estimated time of emergence of this phytoplasma lineage.

The last two examples support the hypothesis of a macroevolutionary pattern of pathogen diversification mediated by host-shifts. Based on available host data, we speculate that shifts in host plants by insect vectors may have driven rapid radiation and diversification of phytoplasmas. As suggested by Poinar (2014), when a host shift results in a highly specialized association this will lead to a less stable pathosystem, whereas less specialized associations lead to increased resilience against abiotic and biotic constraints. In our representative dataset, several diverse lineages (e.g. 16SrI, 16SrII) are associated with many different plant and insect families. Although preliminary analyses of patterns of phytoplasma-plant and phytoplasma-insect associations suggest more stable pathosystems for some groups of phytoplasmas, it was also noted that available records of phytoplasma-host associations are affected by a sampling bias due to under-sampling of interactions that occur in natural habitats (Trivellone, 2019; Trivellone and Flores Garcia, 2019). So far, almost all investigations of phytoplasma-plant-vector interactions have aimed to clarify the epidemiology of phytoplasma-associated diseases in agroecosystems.

### 3.3. Conclusions

More research is needed on phytoplasma-host associations in natural habitats to facilitate robust analyses of the origin of this ancient lineage of pathogens and specific evolutionary patterns related to host shifts, biogeography and other macroevolutionary processes.

A detailed estimate of divergence times in phytoplasmas is relevant for further research aimed to reveal the macro- and microevolutionary processes driving diversification patterns. Our results indicate that the phytoplasma lineage is much older (~640.71 Ma) than previously inferred (180 Ma based on known geological events but in the absence of molecular divergence information). This very ancient origin implies that the lineage that eventually gave rise to phytoplasmas retained a free-living lifestyle or commensal associations with the ancestors of vertebrates or Hexapoda for hundreds of millions of years before acquiring the highly specialized parasitic/plant-pathogenic lifestyle shared by all modern members of the group. This study represents a first attempt to develop a detailed timetree for phytoplasmas using molecular divergence time estimation methods. Further investigations of the interactions of phytoplasmas with their host plants and vectors are needed in order to fill major gaps in knowledge of the long evolutionary history of this group of pathogenic bacteria.

Author contribution statements

CHD suggested and supervised the project. YC and VT contributed equally to compilation and analysis of the data, drafting the manuscript and designing the figures. All authors discussed the results and contributed to the final manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was funded in part by U.S. National Science Foundation grant DEB – 16-39601.

We would like to thank the two anonymous referees for the careful and insightful review of our manuscript. We also thank all the researchers who have provided the sequences on NCBI without whom this work would not be possible.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.106826.

## References

Anderson, C.L., Janssen, T., 2009. Monocots. In: Hedges, S.B., Kumar, S., editors. Timetree of life [Internet]. Oxford University Press (Oxford biology). http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-96062.

Angelini, E., Clair, D., Borgo, M., Bertaccini, A., Boudon-Padieu, E., 2001. Flavescence doree in France and Italy - occurrence of closely related phytoplasma isolates and their near relationships to Palatinate grapevine yellows and an alder yellows phytoplasma. Vitis 40 (2), 79–86.

Angelini, E., Negrisolo, E., Clair, D., Borgo, M., Boudon-Padieu, E., 2003. Phylogenetic relationships among Flavescence doree strains and related phytoplasmas determined by heteroduplex mobility assay and sequence of ribosomal and Nonribosomal DNA. Plant Pathol. 52 (5), 663–672.

Angelini, E., Squizzato, F., Lucchetta, G., Borgo, M., 2004. Detection of a phytoplasma associated with grapevine Flavescence dorée in Clematis vitalba. Eur. J. Plant. Pathol. 110 (2), 193–201.

Arnaud, G., Malembic-Maher, S., Salar, P., Bonnet, P., Maixner, M., Marcone, C., et al., 2007. Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence doree phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. Appl Environ Microbiol. 73 (12), 4001–4010.

Bell, M.A., Lloyd, G.T., 2015. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. Palaeontology. 58 (2), 379–389.

Bertaccini, A., Duduk, B., Paltrinieri, S., Contaldo, N., 2014. Phytoplasmas and phytoplasma diseases: a severe threat to agriculture. Am. J. Plant. Sci. 05 (12), 1763–1788.

Bonfils, J., Schvester, D., 1960. The leafhoppers (Homoptera: Auchenorrhyncha) and their relationship with vineyards in south-western France. Ann. Epiphytes. 325–336.

Brown, D.R., 2010. Phylum XVI. Tenericutes Murray 1984a, 356VP (Effective publication: Murray 1984b, 33). In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, et al., editors. Bergey's Manual of Systematic Bacteriology: Volume Four The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes [Internet]. New York, NY: Springer New York. p. 567–723. https://doi.org/10.1007/978-0-387-68572-4_5.

Brown, D.R.B., Bradbury, J.M., Johansson, K.E., 2010. In: Ludwig, W., Whitman, W., Hedlund, B.P., Paster, B.J., Staley, J.T., Ward, N. (Eds.), 2nd ed. Bergey's Manual of Systematic Bacteriology: Volume 4: The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes [Internet] 4. Springer-VerlagBergey's Manual of Systematic Bacteriology, New York, pp. 687. https://www.springer.com/us/book/9780387950426.

Brown, D.R.B., Johansson, J.M., 2010. Genus, I., 1970. Acholeplasma Edward and Freundt 1970. In: Krieg NR, Ludwig W, Whitman W, Hedlund BP, Paster BJ, Staley JT, et al., editors. Bergey's Manual of Systematic Bacteriology: Volume 4: The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes [Internet]. 2nd ed. New York: Springer-Verlag. p. 688–96. (Bergey's Manual of Systematic Bacteriology; vol. 4). https://www.springer.com/us/book/9780387950426.

Caudwell, A., 1957. Deux années d'études sur la flavescence dorée nouvelle maladie grave de la vigne. Institut National de la Recherche agronomique, Paris.

Caudwell, A., 1964. Identification d'une nouvelle maladie à virus de la vigne, la Flavescence dorée: étude des phénomènes de localisation des symptômes et de rétablissement. Vol. 15. Paris. 193 p.

Caudwell, A., Kuszala, C., Bachelier, J.C., Larrue, J., 1970. Transmission de la Flavescence dorée de la vigne aux plantes herbacées par l'allongement du temps d'utilisation de la cicadelle Scaphoideus littoralis Ball et l'étude de sa survie sur un grand nombre d'espèces végétales. Ann Epiphytes. 2, 415–428.

Chang, S.H., Tan, C.M., Wu, C.-T., Lin, T.-H., Jiang, S.-Y., Liu, R.-C., et al., 2018. Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. J. Exp. Bot. 69 (22), 5389–5401.

Chriki-Adeeb, R., Chriki, A., 2016. Estimating divergence times and substitution rates in rhizobia. Evol. Bioinforma Online. 12, 87–97.

Chuche, J., Thiéry, D., 2014. Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: a review. Agron Sustain. Dev. 34, 381–403.

Daire, X., Clair, D., Reinert, W., Boudon-Padieu, E., 1997. Detection and differentiation of grapevine yellows phytoplasmas belonging to the elm yellows group and to the stolbur subgroup by PCR amplification of non-ribosomal DNA. Eur. J. Plant Pathol. 103 (6), 507–514.

Davis, J.J., Xia, F., Overbeek, R.A., Olsen, G.J., 2013. Genomes of the class Erysipelotrichia clarify the firmicute origin of the class Mollicutes. Int. J. Syst. Evol/Microbiol. 63 (7), 2727–2741.

Doi, Y., Teranaka, M., Yora, K., Asuyama, H., 1967. Mycoplasma- or PLT group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows, or paulownia witches' broom. Jpn J. Phytopathol. 33 (4), 259–266.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C.J., Yang, Z., 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25 (22), 2939–2950.

Edgecombe, G.D., Legg, D.A., 2014. Origins and early evolution of arthropods. Smith A, editor. Palaeontology. 2014; 57(3), 457–68.

Federhen, S., 2012. The NCBI Taxonomy database. Nucleic Acids Res. 2012; 40 (Database issue): D136–43. http://www.ncbi.nlm.nih.gov/taxonomy (accessed 28 July 2019).

Fletcher, J., Melcher, U., Wayadande, A., 2006. The Phytopathogenic Spiroplasmas. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., editors. The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria [Internet]. New York, NY: Springer US; 2006. p. 905–47. https://doi.org/10.1007/0-387-30744-3_30.

Giribet, G., Edgecombe, G.D., 2019. The phylogeny and evolutionary history of arthropods. Curr. Biol. 29 (12), R592–R602.

Gruen, D.S., Wolfe, J.M., Fournier, G.P., 2019. Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. BMC Evol. Biol. 19 (1), 34.

Gundersen, D.E., Lee, I.M., Rehner, S.A., Davis, R.E., Kingsbury, D.T., 1994. Phylogeny of mycoplasmalike organisms (phytoplasmas): a basis for their classification. J. Bacteriol. 176 (17), 5244–5254.

Gupta, R.S., Sawnani, S., Adeolu, M., Alnajar, S., Oren, A., 2018. Phylogenetic framework for the phylum Tenericutes based on genome sequence data: proposal for the creation of a new order Mycoplasmoidales ord. nov., containing two new families Mycoplasmoidaceae fam. nov. and Metamycoplasmataceae fam. nov. harbouring Eperythrozoon, Ureaplasma and five novel genera. Antonie Van Leeuwenhoek. 111 (9), 1583–1630.

Harrison, N.A., Gundersen-Rindal, D., Davis, R.E., May, M., Brown, D.R., 2015. Candidatus Phytoplasma. In: Bergey's Manual of Systematics of Archaea and Bacteria [Internet]. American Cancer Society; 2015. p. 1–38. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118960608.gbm01259.pub2.

Hogenhout, S.A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H.N., Namba, S., 2008. Phytoplasmas: bacteria that manipulate plants and insects. Mol. Plant. Pathol. 9 (4), 403–423.

IRPCM, 2004. Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group. 'Candidatus Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int. J. Syst. Evol. Microbiol. 2004; 54(4): 1243–55.

Jensen, J.S., 2017. Mycoplasma and Ureaplasma. In: Infectious Diseases [Internet]. Elsevier; 2017. p. 1660-1665.e2. https://linkinghub.elsevier.com/retrieve/pii/B9780702062858001866.

Johansson, K.-E., Heldtander, M.U.K., Pettersson, B., 1998. Characterization of

Mycoplasmas by PCR and Sequence Analysis with Universal 16S rDNA Primers. In: Miles, R., Nicholas, R., editors. Mycoplasma Protocols [Internet]. Totowa, NJ: Humana Press; 1998. p. 145–65. (Methods in Molecular BiologyTM). https://doi.org/10.1385/0-89603-525-5:145.

Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., et al., 2018. Phylogenomics and the evolution of hemipteroid insects. Proc. Natl. Acad. Sci. 115 (50), 12775–12780.

Jung, H.-Y., Sawayanagi, T., Kakizawa, S., Nishigawa, H., Wei, W., Oshima, K., et al., 2003. "Candidatus Phytoplasma ziziphi", a novel phytoplasma taxon associated with jujube witches'-broom disease. Int. J. Syst. Evol. Microbiol. 53 (4), 1037–1041.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874.

Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. Timetree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34 (7), 1812–1819.

Kuo, C.-H., Ochman, H., 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol. Direct. 4, 35.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14 (1), 82.

Lee, I.-M., Bottner-Parker, K.D., Zhao, Y., Davis, R.E., Harrison, N.A., 2010. Phylogenetic analysis and delineation of phytoplasmas based on secY gene sequences. Int. J. Syst. Evol. Microbiol. 60 (12), 2887–2897.

Ludwig, W., Klenk, H.-P., 2001. Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In: Boone, D.R., Castenholz, R.W., Garrity, G.M., editors. Bergey's Manual of Systematic Bacteriology: Volume One : The Archaea and the Deeply Branching and Phototrophic Bacteria [Internet]. New York, NY: Springer New York; 2001. p. 49–65. https://doi.org/10.1007/978-0-387-21609-6_8.

MacLean, A.M., Sugio, A., Makarova, O.V., Findlay, K.C., Grieve, V.M., Tóth, R., et al., 2011. Phytoplasma effector SAP54 induces indeterminate leaf-like flower development in arabidopsis plants. Plant Physiol. 157 (2), 831–841.

Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., et al., 2014. Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. Plant. J. Cell Mol. Biol. 78 (4), 541–554.

Malembic-Maher, S., Desque, D., Khalil, D., Salar, P., Danet, J.L., Bergey, B., et al. When a Palearctic bacterium meets a Nearctic insect vector: Genetic and ecological insights into the emergence of the grapevine Candidatus Phytoplasma vitis and Ca. P. vitialni epidemics in Europe. In: Future IPM 30 towards a sustainable agriculture [Internet]. Riva del Garda (VR, Italy); 2017. p. 211–3. https://esa.confex.com/esa/ice2016/meetingapp.cgi/Paper/93532.

Maniloff, J., 2002. Phylogeny and Evolution. In: Razin S, Herrmann R, editors. Molecular Biology and Pathogenicity of Mycoplasmas [Internet]. Boston, MA: Springer US; 2002. p. 31–43. https://doi.org/10.1007/0-306-47606-1_2.

Malembic-Maher, S., Salar, P., Filippin, L., Carle, P., Angelini, E., Foissac, X., 2011. Genetic diversity of European phytoplasmas of the 16SrV taxonomic group and proposal of '*Candidatus* Phytoplasma rubi'. International Journal of Systematic and Evolutionary Microbiology 61 (9), 2129–2134. https://doi.org/10.1099/ijs.0.025411-0.

Marin, J., Battistuzzi, F.U., Brown, A.C., Hedges, S.B., 2016. The Timetree of prokaryotes: new insights into their evolution and speciation. Mol. Biol. Evol. 34 (2), 437–446.

Martini, M., Lee, I.-M., Bottner, K.D., Zhao, Y., Botti, S., Bertaccini, A., et al., 2007. Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. Int. J. Syst. Evol. Microbiol. 57 (9), 2037–2051.

Martini, M., Marcone, C., Lee, I.-M., Firrao, G.. The Family Acholeplasmataceae (Including Phytoplasmas). In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes: Firmicutes and Tenericutes [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 469–504. https://doi.org/10.1007/978-3-642-30120-9_387.

Mello, B., 2018. Estimating TimeTrees with MEGA and the TimeTree Resource. Mol Biol Evol. 35 (9), 2334–2342.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., et al., 2018. The timescale of early land plant evolution. Proc. Natl. Acad Sci. 115 (10), E2274–E2283.

Murray, R.G.E., 1984. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. List no. 15. Int. J. Syst. Bacteriol. 34, 355–357.

Namba, S., Oyaizu, H., Kato, S., Iwanami, S., Tsuchizaki, T., 1993. Phylogenetic diversity of phytopathogenic mycoplasmalike organisms. Int. J. Syst. Bacteriol. 43 (3), 461–467.

Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. 96 (22), 12638–12643.

Ogawa, Y., Ooka, T., Shi, F., Ogura, Y., Nakayama, K., Hayashi, T., et al., 2011. The genome of Erysipelothrix rhusiopathiae, the causative agent of swine erysipelas, reveals new insights into the evolution of firmicutes and the organism's intracellular adaptations. J. Bacteriol. 193 (12), 2959–2971.

Ollat, N., Bordenave, L., Tandonnet, J.P., Boursiquot, J.M., Marguerit, E., 2016. Grapevine rootstocks: origins and perspectives. In: Acta Horticulturae [Internet]. International Society for Horticultural Science (ISHS), Leuven, Belgium; 2016. p. 11–22. https://doi.org/10.17660/ActaHortic.2016.1136.2.

Papura, D., Burban, C., van Helden, M., Giresse, X., Nusillard, B., Guillemaud, T., et al., 2012. Microsatellite and Mitochondrial Data Provide Evidence for a Single Major Introduction for the Neartic Leafhopper Scaphoideus titanus in Europe. PLoS One [Internet]. 7 (5).

Parte, A.C., 2018. LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. Int. J. Syst. Evol. Microbiol. 68 (6), 1825–1829.

Percy, D.M., Crampton-Platt, A., Sveinsson, S., Lemmon, A.R., Lemmon, E.M., Ouvrard, D., et al., 2018. Resolving the psyllid tree of life: phylogenomic analyses of the superfamily Psylloidea (Hemiptera): Resolving the psyllid tree of life. Syst. Entomol. 43 (4), 762–776.

Pérez-López, E., Olivier, C.Y., Luna-Rodríguez, M., Dumonceaux, T.J., 2016. Phytoplasma classification and phylogeny based on in silico and in vitro RFLP analysis of cpn60 universal target sequences. Int. J. Syst. Evol. Microbiol. 66 (12), 5600.

Pettersson, B., Uhlén, M., Johansson, K.E., 1996. Phylogeny of some mycoplasmas from ruminants based on 16S rRNA sequences and definition of a new cluster within the hominis group. Int. J. Syst. Bacteriol. 46 (4), 1093–1098.

Plavec, J., Budinšćak, Ž., Križanac, I., Škorić, D., Foissac, X., Šeruga, Musić M., 2019. Multilocus sequence typing reveals the presence of three distinct flavescence dorée phytoplasma genetic clusters in Croatian vineyards. Plant Pathol. 68 (1), 18–30.

Poinar G. Evolutionary History of Terrestrial Pathogens and Endoparasites as Revealed in Fossils and Subfossils [Internet]. Advances in Biology. 2014 https://www.hindawi.com/journals/ab/2014/181353/.

R Core Team, 2019. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/.

Razin, S., 2006. The Genus Mycoplasma and Related Genera (Class Mollicutes). In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E., editors. The Prokaryotes [Internet]. New York, NY: Springer US; 2006. p. 836–904. http://link.springer.com/10.1007/0-387-30744-3_29.

Razin, S., Yogev, D., Naot, Y., 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Mol. Biol. Rev. 62 (4), 1094–1156.

Rivera-Tapia, J.A., Cedillo-Ramírez, M.L., Juárez, C.G., 2002. Some biological features of Mollicutes. Rev. Latinoam Microbiol. 44 (2), 53–57.

Sanglas, A., Albarral, V., Farfán, M., Lorén, J.G., Fusté, M.C., 2017. Evolutionary roots and diversification of the genus aeromonas. Front. Microbiol. 8, 127.

Schvester, D., Carle, P., Moutous, G., 1962. Sur la transmission de la Flavescence dorée des vignes par une cicadelle. CR Acad Agric. 47, 1021–1024.

Sheridan, P.P., Freeman, K.H., Brenchley, J.E., 2003. Estimated minimal divergence times of the major bacterial and archaeal phyla. Geomicrobiol. J. 20 (1), 1–14.

Sjöstrand, J., Tofigh, A., Daubin, V., Arvestad, L., Sennblad, B., Lagergren, J., 2014. A Bayesian method for analyzing lateral gene transfer. Syst. Biol. 63 (3), 409–420.

Skennerton, C.T., Haroon, M.F., Briegel, A., Shi, J., Jensen, G.J., Tyson, G.W., et al., 2016. Phylogenomic analysis of *Candidatus* 'Izimaplasma' species: free-living representatives from a Tenericutes clade found in methane seeps. ISME J. 10 (11), 2679–2692.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinforma Oxf Engl. 30 (9), 1312–1313.

Tamura, K., Tao, Q., Kumar, S., 2018. Theoretical foundation of the reltime method for estimating divergence times from variable evolutionary rates. Mol. Biol. Evol. 35 (7), 1770–1782.

Tomkins, M., Kliot, A., Marée, A.F., Hogenhout, S.A., 2018. A multi-layered mechanistic modelling approach to understand how effector genes extend beyond phytoplasma to modulate plant hosts, insect vectors and the environment. Curr. Opin. Plant Biol. 44, 39–48.

Trivellone, V., 2019. An online global database of Hemiptera-Phytoplasma-Plant biological interactions. Biodivers Data J [Internet]. 7.

Trivellone, V., Flores Garcia, C.O., 2019. Network analyses of a global Hemiptera-phytoplasma-plant biological interactions database. Phytopathogenic Mollicutes 9 (1), 35.

Tully, J.G., 1996. Mollicute-host interrelationships: Current concepts and diagnostic implications. In: Tully, J.G., Razin, S. (Eds.), Molecular and Diagnostic ProceduresinMycoplasmology: Diagnostic Procedures. Academic PRess, San Diego, CA, United Sates, pp. 1–21.

Volokhov, D.V., Neverov, A.A., George, J., Kong, H., Liu, S.X., Anderson, C., et al., 2007. Genetic analysis of housekeeping genes of members of the genus Acholeplasma: phylogeny and complementary molecular markers to the 16S rRNA gene. Mol. Phylogenet. Evol. 44 (2), 699–710.

Waites, K.B., 2016. Medically Relevant Mycoplasmas and Ureaplasmas. In: Reference Module in Biomedical Sciences [Internet]. Elsevier; 2016. http://www.sciencedirect.com/science/article/pii/B9780128012383022996.

Wei, W., Davis, R.E., Lee, I.-M., Zhao, Y., 2007. Computer-simulated RFLP analysis of 16S rRNA genes: identification of ten new phytoplasma groups. Int. J. Syst. Evol. Microbiol. 57 (8), 1855–1867.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., et al., 1989. A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. 171 (12), 6455–6467.

Wolf, M., Müller, T., Dandekar, T., Pollack, J.D., 2004. Phylogeny of Firmicutes with special reference to Mycoplasma (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Int. J. Syst. Evol. Microbiol. 54 (3), 871–875.

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., et al., 2017. Evolution of rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol. Biol. Evol. 34 (2), 262–281.

Yutin, N., Galperin, M.Y., 2013. A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced clostridia. Environ. Microbiol. 15 (10), 2631–2641.

Zhao, Y., Wei, W., Davis, R.E., Lee, I.M., 2009. Recent advances in 16S rRNA gene-based phytoplasma differentiation, classification and taxonomy. In: Weintraub, P.G., Jones, P. (Eds.), Phytoplasmas: genomes, plant hosts and vectors [Internet]. CABI, Wallingford, pp. 64–92 http://www.cabi.org/cabebooks/ebook/20093353128.