Revisiting the taxonomy of *Allorhizobium vitis* (i.e. *Agrobacterium vitis*) using genomics - emended description of *All. vitis sensu stricto* and description of *Allorhizobium ampelinum* sp. nov.

Nemanja Kuzmanović[1,*], Enrico Biondi[2], Jörg Overmann[3], Joanna Puławska[4], Susanne Verbarg[3], Kornelia Smalla[1], Florent Lassalle[5,6,*]

[1]Julius Kühn-Institut, Federal Research Centre for Cultivated Plants (JKI), Institute for Epidemiology and Pathogen Diagnostics, Messeweg 11-12, 38104 Braunschweig, Germany

[2]Alma Mater Studiorum - University of Bologna, Viale G. Fanin, 42, 40127 Bologna, Italy

[3]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffenstrasse 7B, 38124 Braunschweig, Germany

[4]Research Institute of Horticulture, ul. Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

[5]Imperial College London, St-Mary's Hospital campus, Department of Infectious Disease Epidemiology, Praed Street, London W2 1NY, UK; Imperial College London, St-Mary's Hospital campus, MRC Centre for Global Infectious Disease Analysis, Praed Street, London W2 1NY, United Kingdom

[6]Wellcome Sanger Institute, Pathogens and Microbes Programme, Wellcome Genome Campus, Hinxton, Saffron Walden, CB10 1RQ, United Kingdom

*Corresponding authors.
Contact: kuzmanovic1306@gmail.com, nemanja.kuzmanovic@julius-kuehn.de (N. Kuzmanović); f.lassalle@imperial.ac.uk (F. Lassalle)

Abstract

Allorhizobium vitis (formerly named *Agrobacterium vitis* or *Agrobacterium* biovar 3) is the primary causative agent of crown gall disease of grapevine worldwide. Whole-genome sequence comparisons and phylogenomic analysis of various *All. vitis* strains clearly indicated that *All. vitis* is not a single species, but represents a species complex composed of at least four genomic species. Thus, we amended the description of *All. vitis* which now refers to a restricted group of strains within the *All. vitis* complex (i.e. *All. vitis sensu stricto*) and proposed a description of a novel species *All. ampelinum* sp. nov. The type strain of *All. vitis sensu stricto* remains the existing type strain of *All. vitis*, K309^T (= NCPPB 3554^T =HAMBI 1817^T = ATCC 49767^T = CIP 105853^T = ICMP 10752^T = IFO 15140^T = JCM 21033^T = LMG 8750^T = NBRC 15140^T). The type strain of *All. ampelinum* sp. nov. is S4^T (= DSM 112012^T = ATCC BAA-846^T). This genome-based classification was supported by differentiation of strains based on a MALDI-TOF MS analysis. We also identified gene clusters specific for *All. vitis* species complex, *All. vitis sensu stricto* and *All. ampelinum,* and attempted to predict their function and their role in ecological diversification of these clades, some of which were experimentally validated. Functions of *All. vitis* species complex-specific genes convergently indicate a role in adaptation to different stresses, including exposure to aromatic compounds. Similarly, *All vitis sensu stricto*-specific genes also confer the ability to degrade 4-hydroxyphenylacetate and a putative compound related to gentisic acid, while *All. ampelinum*-specific genes have putative functions related to polyamine metabolism and nickel assimilation. This suggests that these species have differentiated ecologies, each relying on specialized nutrient consumption or toxic compound degradation to adapt to their respective niche. Moreover, our genome-based analysis indicated that *Allorhizobium* and the "*R. aggregatum* complex" represent separate genera of the family *Rhizobiaceae*.

Keywords: Rhizobiaceae, taxonomy, plant pathogenic bacteria, clade-specific genes; ecological specialization; pangenome analysis

INTRODUCTION

Allorhizobium vitis (formerly named *Agrobacterium vitis* or *Agrobacterium* biovar 3) is a bacterium primarily known as a plant pathogen causing crown gall disease of grapevine (*Vitis vinifera*) (Kuzmanović et al., 2018). This economically important plant disease may cause serious losses in nurseries and vineyards. *All. vitis* is widely distributed pathogen, detected in almost all grapevine growing regions throughout the world. This bacterium seems to be associated almost exclusively with grapevine. It has been isolated from crown gall tumors, xylem sap, roots, rhizosphere, non-rhizosphere soil of infected vineyards, decaying grape roots and canes in soil, but also from the phyllosphere of grapevine plants [reviewed in (Kuzmanović et al., 2018)]. In one exceptional case, *All. vitis* was isolated from galls on the roots of kiwi in Japan (Sawada & Ieki, 1992).

All. vitis is an aerobic, non-spore-forming, Gram-negative, rod-shaped bacterium with peritrichous flagella (Young et al., 2005). It is a member of the alphaproteobacterial family *Rhizobiaceae*, together with other genera hosting tumor-inducing plant pathogens, including *Agrobacterium* and *Rhizobium*. With time, the taxonomy of *All. vitis* has undergone various changes. Tumorigenic strains associated with crown gall of grapevine were initially defined as an atypical group that could neither be classified as *Agrobacterium* biovar 1 (i.e., *Agrobacterium tumefaciens* complex) nor as biovar 2 (i.e. *Rhizobium rhizogenes*) (Panagopoulos & Psallidas, 1973). Afterwards, several studies classified these atypical strains as *Agrobacterium* biovar 3 (biotype 3), based on their biochemical and physiological characteristics (Kerr & Panagopoulos, 1977, Panagopoulos et al., 1978, Süle, 1978). Serological analysis using monoclonal antibodies also allowed differentiation of *Agrobacterium* biovar 3 strains (Bishop et al., 1989). Polyphasic characterization involving DNA-DNA hybridization (DDH), and phenotypic and serological tests clearly showed that *Agrobacterium* biovar 3 strains represent a separate species for which the name *Agrobacterium vitis* was proposed (Ophel & Kerr, 1990). However, multi-locus sequence analysis (MLSA) suggested that *A. vitis* is phylogenetically distinct from the newly redescribed genus *Agrobacterium*, and prompted the transfer of this species to the revived genus *Allorhizobium* (Mousavi et al., 2014, Mousavi et al., 2015).

The genus *Allorhizobium* was created by de Lajudie et al. (1998) and initially included single species *Allorhizobium undicola*. Afterwards Young et al. (2001) proposed reclassification of *All. undicola* and inclusion into the genus *Rhizobium*, while Costechareyre et al. (2010) suggested that this species might belong to the genus *Agrobacterium*. However, these studies employed single gene phylogenies insufficient to support such taxonomic revisions. The authenticity of the genus *Allorhizobium* and the clustering of *All. vitis* within this genus was unequivocally confirmed by genome-wide phylogenies (Ormeño-Orrillo et al., 2015, Hördt et al., 2020). Moreover, distinctiveness of *All. vitis* in respect to the genus *Agrobacterium* was further supported by their genome organization. Genome organization of the genus *Agrobacterium* is characterized by the presence of a circular chromosome and a secondary linear chromid (Ramírez-Bahena et al., 2014, Slater et al., 2009). Chromids are defined as large non-dispensable plasmids carrying essential functions (Harrison et al., 2010). In contrast to *Agrobacterium*, the *All. vitis* strains carry two circular chromosomes (Jumas-Bilak et al. 1998; Tanaka et al. 2006; Slater et al. 2009). However, the smaller chromosome (named chromosome II) was later classified as a chromid in the fully sequenced

strain *All. vitis* S4 (Harrison et al., 2010). Additionally, genomes of *All. vitis* and other agrobacteria include a variable number of plasmids.

In recent years, genomics has significantly impacted the taxonomy of bacteria, leading to the revisions in classification of different bacterial taxa. In particular, a novel genomics-based taxonomy primarily involves calculation of various overall genome relatedness indices (OGRIs) and estimation of genome-based phylogenies (Konstantinidis & Tiedje, 2005, Chun & Rainey, 2014, Parks et al., 2018), largely replacing traditionally used 16S rRNA gene phylogeny and DDH (Wayne et al., 1987, Stackebrandt & Goebel, 1994). Genomic information were also highly recommended as essential for the description of new rhizobial and agrobacterial taxa (de Lajudie et al., 2019). In addition, it has been recommended that some functions and phenotypic characters may not be considered for taxonomic classification. This particularly applies to the tumor-inducing ability of agrobacteria, which is mainly determined by a dispensable tumor-inducing (Ti) plasmid.

Information on genetic diversity and relatedness of strains responsible for crown gall disease outbreaks provide important insights into the epidemiology, ecology and evolution of the pathogen. Numerous studies indicated that *All. vitis* strains are genetically very diverse [reviewed in (Kuzmanović et al., 2018)]. Thus, our previous study analyzing a representative collection of *All. vitis* strains originating from several European countries, Africa, North America, and Australia using MLSA indicated a high genetic diversity between strains, which clustered into four main phylogenetic groups (Kuzmanović et al., 2015). These data suggested that *All. vitis* might not be a homogenous species, but a species complex comprising several genomic species. This warrants additional investigation of the diversification and evolution of *All. vitis* towards further elucidation of the taxonomy of this group.

In this work, we selected representative strains belonging predominantly to the two most frequent phylogenetic groups identified in our previous study (Kuzmanović et al., 2015), hosting the well-studied *All. vitis* type strain K309^T and fully sequenced strain S4. We obtained draft genome sequences for 11 additional strains and performed comparative genome analyses to reveal phylogeny, diversification and specific features of these groups. In parallel, we investigated phenotypic features of selected strains. The combination of these approaches allows us to revise the taxonomy within this group, notably by proposing the new species *All. ampelinum*.

MATERIALS AND METHODS

Allorhizobium vitis strains

All. vitis strains used in this study were isolated from crown gall tumors on grapevine originating from different geographical areas (Table 1). These strains were predominantly representatives of the two main phylogenetic groups (C and D) delineated in our previous study (Kuzmanović et al., 2015).

Biochemical tests

All. vitis strains were phenotypically characterized using API and Biolog tests. The API 20NE kit was used according to manufacturer's instructions (bioMérieux). Utilization of sole carbon sources was tested with Biolog GEN III microplates using protocol A, according to the instructions of the manufacturer (Biolog, Inc., Hayward, CA, USA).

The metabolism of 4-hydroxyphenylacetic acid (p-hydroxyphenylacetic acid; Acros Organics, Product code: 121710250) and gentisic acid (2,5-dihydroxybenzoic acid; Merck, Product Number: 841745) was performed in AT minimal medium (Tempé et al., 1977, Morton & Fuqua, 2012) supplemented with yeast extract (0.1 g/L), bromthymol blue (2.5 ml/L of 1 % [w/v] solution made in 50% ethanol), and the tested compound (1 g/L). Hydroxyphenylacetic and gentisic acid were added as filter-sterilized 1% aqueous solutions. Bacterial growth and color change of the medium were monitored during one week of incubation at 28°C and constant shaking (200 rpm/min). Metabolism of L(+)-tartaric acid, involving production of alkali from this compound, was tested as described before (Kerr & Panagopoulos, 1977).

MALDI-TOF Mass Spectrometry analysis

Sample preparation for MALDI-TOF mass spectrometry (MS) was carried out according to Protocol 3 described by Schumann and Maier (2014). Instrument settings for the measurement were as described previously by Tóth et al. (2008). The dendrogram was created using the MALDI Biotyper Compass Explorer software (Bruker, Version 4.1.90).

DNA extraction

For whole genome sequencing, genomic DNA was extracted from bacterial strains grown on King's medium B (King et al. 1954) at 28°C for 24 h using NucleoSpin Microbial DNA kit (Macherey-Nagel, Germany). The quality of the genomic DNA was assessed by electrophoresis in 0.8% agarose gel.

Genome sequencing

Draft whole-genome sequences were obtained for 11 *All. vitis* strains (Table 1). DNA libraries were obtained with Nextera XT DNA Library Prep Kit (Illumina, USA). Paired-end sequencing (2 × 300 bp) was performed on an Illumina MiSeq platform generating 487,883-2,309,377 paired reads per genome. Trimming and quality filtering of raw reads were conducted using Trimmomatic (Galaxy Version 0.36.5) (Bolger et al., 2014) implemented on the Galaxy Web server (Afgan et al., 2018). The read quality was assessed with FastQC (Galaxy Version 0.72+galaxy1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). In order to achieve higher coverage for strains Av2, IPV-BO 1861-5, KFB239 and KFB 264, additional paired-end sequencing (2 × 150 bp) was performed using an Illumina NextSeq 500 platform generating 1,037,619-1,443,575 paired reads. Demultiplexing and adapter clipping was done using bcl2fastq(2) conversion software (Illumina, USA).

Genome assembly and annotation

De novo genome assemblies were performed using the SPAdes genome assembler (Bankevich et al., 2012) (Galaxy Version 3.12.0+galaxy1). For genomes sequenced on the MiSeq and NextSeq platforms, both sets of reads were used for assembly. The genome sequences were annotated using Prokka (Galaxy Version 1.13) (Seemann, 2014) and NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) (Tatusova et al., 2016). Prokka Version 1.14.6 was used to annotate genomes as a part of Pantagruel pipeline (task 0; see below).

Core- and pan-genome phylogenomic analyses

For phylogenomic analysis, whole genome sequences of 103 *Rhizobiaceae* strains were used, including 48 strains of *All. vitis* complex (Tables 1 and S1). In this dataset (dataset 1), all *All. vitis* genome sequences available in the GenBank and those sequenced in this study were included. The second dataset (dataset 2) included reference *Rhizobiaceae* species and only *All. vitis* genomes sequenced in this study, as well as high quality genomes sequenced, deposited and published by other groups (Table 1; Table S1).

GenBank files generated by Prokka were used as an input. Homologous gene clusters were computed using bidirectional best-hit (BDBH), Clusters of Orthologous Groups-triangles (COGtriangles), and OrthoMCL (Markov Clustering of orthologs, OMCL) algorithms by running the get_homologues.pl script implemented into GET_HOMOLOGUES software package Version 10032020 (Contreras-Moreira & Vinuesa, 2013). A stringent 90% coverage cut-off for BLASTP alignments (-C 90) was imposed. A consensus core-genome was computed as the intersection of the clusters computed by the BDBH, COG-triangles and OMCL algorithms by employing script compare_clusters.pl (-t "number of genomes"). The resulting core-genome clusters were processed with the GET_PHYLOMARKERS software package Version 2.2.8.1_16Jul2019 (Vinuesa et al., 2018) by using a pipeline for DNA-based phylogenies (-R 1 -t DNA). This software was designed to select reliable phylogenetic markers and exclude sequences with the evidence of recombination, and those yielding anomalous and poorly resolved gene trees. The selected DNA sequences were codon aligned and concatenated, producing a supermatrix alignment that were used for phylogenetic analysis. A maximum likelihood (ML) phylogeny was estimated under the best-fitting substitution model using IQ-TREE v1.6.10 (Nguyen et al., 2015) implemented in the GET_PHYLOMARKERS package.

For pan-genome phylogeny, Dataset 2 (see above) was used. A consensus pan-genome and pan-genome (presence-absence) matrices were computed from the COGtriangles and OMCL clusters (see above) by compare_clusters.pl script (-t 0, -m) from the GET_HOMOLOGUES software package. The ML pangenome phylogeny was estimated from the pan-genome matrix (PGM) by using estimate_pangenome_phylogenies.sh script, bundled with GET_PHYLOMARKERS. Twenty-five independent IQ-TREE searches fitting binary models were launched and the best fit was retained. Another phylogeny based on the PGM was inferred using parsimony with the script compare_clusters.pl (option –T) from the GET_HOMOLOGUES package.

Overall genome relatedness indices

To differentiate among the strains studied, different OGRIs were computed. For species delimitation, the average nucleotide identity (ANI) (Richter & Rossello-Mora, 2009, Goris et al., 2007) among the strains was calculated using PyANI program Version 0.2.9, with scripts employing BLAST+ (ANIb) and MUMmer (ANIm) to align the input sequences (https://github.com/widdowquinn/pyani). Additionally, ANI values were calculated by OrthoANIu Version 1.2 (calculates orthologous ANI using USEARCH algorithm) (Yoon et al., 2017) and FastANI Version 1.2 (estimates ANI using Mashmap as its MinHash based alignment-free sequence mapping engine) (Jain et al., 2018) tools. The *in silico* DNA-DNA hybridization (isDDH) values were calculated using the Genome-to-Genome Distance Calculator (GGDC 2.1; http://ggdc.dsmz.de/distcalc2.php) by employing the recommended BLAST+ alignment and formula 2 (identities/HSP length) (Meier-Kolthoff et al., 2013).

For genus delimitation, average amino acid identity (AAI) (Goris et al., 2007, Konstantinidis & Tiedje, 2005, Konstantinidis et al., 2017) values were calculated with CompareM Version 0.0.23 (https://github.com/dparks1134/CompareM) by using default options. Genome-wide average nucleotide identity (gANI) and alignment fraction (AF) values were obtained by the ANIcalculator Version 1.0 (Varghese et al., 2015). This gANI calculation involves comparison of orthologous protein-coding genes identified as bidirectional best hits (BBHs), at the nucleotide level. AF is calculated as a fraction of the sum of the lengths of BBH genes divided by the sum of the lengths of all genes in a genome (Varghese et al., 2015). For gANI/AF analysis, genomes annotated by Prodigal Version 2.6.3 were used. Conveniently, prodigal does not predict RNA genes (tRNA or rRNA) that can artificially inflate the gANI/AF. The percentage of conserved proteins (POCP) (Qin et al., 2014) values were calculated using the OMCL algorithm (Li et al., 2003) through GET_HOMOLOGUES software package (option -P) (Contreras-Moreira & Vinuesa, 2013).

Genome composition analyses of *All. vitis* complex strains

Homologous gene clusters of 14 *All. vitis* complex strains studied were computed using COGtriangles and OMCL algorithms, followed by computation of consensus pan-genome clusters and pan-genome matrix, as described above. The pan-genome clusters were classified into core, soft core, cloud and shell compartments (Koonin & Wolf, 2008) by auxiliary script parse_pangenome_matrix.pl (option -s) of GET_HOMOLOGUES software package (Contreras-Moreira & Vinuesa, 2013). Accessory genes include both shell and cloud compartments.

Identification of clade-specific genes

Further pangenome analyses were conducted using the Pantagruel pipeline under the default settings as described previously (Lassalle et al., 2019, Lassalle et al., 2020) and on the program webpage http://github.com/flass/pantagruel/. Because of computationally highly intensive tasks, the dataset analyzed was limited to the *Allorhizobium* genus and *Rhizobium aggregatum* complex (total of 28 strains). Shortly, the Pantagruel pipeline was used to infer evolutionary events like gene duplication, horizontal transfer and loss, for each

gene family of the 28-species pangenome with at least four sequences by reconciling the topology of gene trees with the reference species tree (core genome tree) in a probabilistic framework. Thus, genes were classified into orthologous clusters based on the gain/loss scenarios, from which clade-specific gene sets were identified. Genes specific for a particular clade (species or genus) were identified from the gene_count_matrix output of Pantagruel task 08, using the R script get_clade_specific_genes.r available in Pantagruel.

Moreover, the script parse_pangenome_matrix.pl of GET_HOMOLOGUES software package was employed for the identification of species-specific gene families (option -g) from pan-genome matrix outputted by GET_HOMOLOGUES software package. The same script was used for plotting the species-specific genes on a linearized genetic map of a reference genome selected from that species (options –g and –p) (see also GET_HOMOLOGUES manual: http://eead-csic-compbio.github.io/get_homologues/manual/). As input, pangenome clusters were computed from the same dataset as the one used for Pantagruel analysis (28 strains of *Allorhizobium* and *Rhizobium aggregatum* complex). In addition, the pangenome dataset generated from 14 strains of *All. vitis* complex strains (Table 1) (see "Genome composition analyses of *All. vitis* complex" above) was analyzed.

Functional annotation of proteins encoded by each gene family clustered by Pantagruel was conducted by the InterProScan software package (Jones et al., 2014) implemented in the Pantagruel pipeline (Task 4). Additionally, annotation of particular sequences of interest and metabolic pathway prediction were performed using BlastKOALA and GhostKOALA (Kanehisa et al., 2016b). Protein sequences analyzed were subjected to Pfam domain searches (database release 32.0, September 2018, 17929 entries) (El-Gebali et al., 2018). Metabolic pathway prediction was performed using KEGG (Kanehisa et al., 2016a) and MetaCyc (Caspi et al., 2013) databases. Over-representation of Gene Ontology functional annotation terms in clade-specific genes with respect to the clade's core genomes was tested using the R script clade_specific_genes_GOterm_enrichment_test.r available in Pantagruel, using the enrichment test implemented in the topGO R package (Alexa et al., 2006) with the 'weight01' algorithm and 'Fisher' statistics (tests with p-values lower than 0.1 were deemed significant).

The NCBI BLASTN and BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi), as well as BLAST search tool of KEGG database (Kanehisa et al., 2016a), were used for sequence comparisons at the nucleotide and amino acid levels, respectively.

Accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers listed in Table 1. The versions described in this paper are first versions.

RESULTS

***Allorhizobium vitis* genome sequencing**

In this study, draft genome sequences of 11 *All. vitis* strains were obtained using the Illumina platform. Coverage of genomes obtained in this study ranged from 65-96 fold. Some

basic genome assembly statistics are summarized in Table 1. The total size of the draft genome sequences ranged from 5.67 to 6.52 Mb, with a GC content ranging 57.5-57.6% (Table 1), which was similar to the genomes of other *All. vitis* strains sequenced so far (Table S1).

Core- and pan-genome phylogenies

Core-genome phylogeny was inferred for 48 *All. vitis* strains available in the Genbank (Dataset 1), as well as for a limited set of *All. vitis* genomes sequenced in this study and those published elsewhere (Dataset 2). Both datasets included reference Rhizobiaceae members. The high stringency consensus core-genome contained 515 (Dataset 1; Fig. S1) and 555 (Dataset 2, Fig. 1) homologous gene clusters. Phylogenomic trees were inferred from 302 (Dataset 1, Fig. S1) and 344 (Dataset 2, Fig. 1) top markers that were selected using GET_PHYLOMARKERS software. The ML core-genome phylogeny confirmed the independent clustering of *Allorhizobium* species (Figs. 1, S1 and S2). *All. vitis* strains formed a separate group ("*All. vitis*" complex clade) within the clade comprising other *Allorhizobium* species (Figs. 1 and S1). *All. vitis* strains formed several well supported sub-clades. The relationships of strains within and among these sub-clades are further discussed below. The clade comprising members of the genus *Allorhizobium* was well separated from a sister clade encompassing organisms belonging to the provisionally named "*Rhizobium aggregatum* complex" by Mousavi et al. (2015). Interestingly, representatives of the genus *Ciceribacter* grouped within "*Rhizobium aggregatum* complex" clade.

The ML pangenome phylogeny was estimated from the consensus (COGtriangles and OMCL clusters) PGM containing 33,396 clusters (Fig. 2; Fig. S3). The pangenome phylogeny (Fig. 2) resolved the same sub-clades of *All. vitis* species complex as core-genome phylogeny (Fig. 1). Furthermore, different *Rhizobiaceae* genera and clades were generally differentiated on pangenome ML tree (Fig. 2). Nevertheless, some inconsistencies were observed. In particular, strain *Neorhizobium* sp. NCHU2750 was more closely related to the representatives of the genus *Agrobacterium*, while *Pararhizobium giardinii* H152^T was grouped with *Ensifer* spp. (Fig. 2; Fig. S3). These inconsistencies were also observed in another phylogeny based on the PGM inferred using parsimony (data not shown).

Species diversity of the *Allorhizobium vitis* complex

Overall genome relatedness indices computed in this study (Table S2 and S3), indicated that *All. vitis* is not a single species, but composed of at least four separate species. These species largely corresponded to the subclades of *All. vitis* determined by our phylogenomic analysis (Fig. S1). The only ambiguity was in the taxonomic relationship of subclades C and D. In fact, ANIb values between members of the subclades C and D were in the range 94.5-95 %, which is close to the threshold for species delimitation (Richter & Rossello-Mora, 2009). However, the subject of our present research were primarily the representatives of subclades A and B, which were sequenced in this study (Figs. 1 and S1).

The first species, designated as *All. vitis sensu stricto*, corresponds to the sub-clade A comprising the type strain of *All. vitis* (strain K309^T) (Fig. 1). Although isDDH values suggested that cluster containing strains K309^T and KFB 253 might belong to a separate species

compared to other strains comprised in this subclade, this was not supported by the other four OGRIs calculated (Table S3). Indeed, it must be noted that isDDH were relatively close to the threshold value of 70 % (65.9-66.4 %). A revised description of *All. vitis sensu stricto* is given below.

The second sub-clade B included eight strains originating from various geographic areas (Table 1; Fig. 1). It included the well-studied strain S4, whose high-quality genome was fully sequenced and finished (Slater et al., 2009). The isDDH value obtained from the comparison of strain KFB 254 with strain IPV-BO 1861-5 was below, but very close to the 70 % threshold value (Table S3). However, other OGRIs unanimously indicated that these strains belong to the same species (Table S3). A description of the novel species corresponding to the subclade B for which the name *Allorhizobium ampelinum* sp. nov. is proposed.

As suggested by phylogenomic analysis, the strain Av2 was more distantly related to sub-clades A and B, which was also supported by OGRIs computed (Table S3; Fig. 1). Nevertheless, it formed a clade with three other strains originating from the USA (Clade D; Fig. S1). Interestingly, these four strains comprised in the clade D were genetically very similar and exhibited >99.8 ANI between each other (Table S2).

Relationship of the genus *Allorhizobium* and "*R. aggregatum* complex"

As indicated by the core-genome phylogeny, the genus *Allorhizobium* was clearly separated from the other representatives of the family *Rhizobiaceae,* including "*R. aggregatum* complex" that formed a well-delineated sister clade to *Allorhizobium* (Figs. 1, S1 and S2). The genome-based comparisons showed a clear divergence between these two clades. In particular, members of the genus *Allorhizobium* shared >74.9% AAI among each other, and 70.79-72.63% AAI with members of "*R. aggregatum* complex" (Table S4). On the other hand, representatives of genera *Shinella*, *Ensifer* and *Pararhizobium* showed 71.46-75.85% AAI similarity between genera. Similarly, representatives of genera *Neorhizobium* and *Pseudorhizobium* showed 72.24-76.18 % AAI similarity between genera. In other words, AAIs suggested that the existing genera *Ensifer*, *Pararhizobium* and *Shinella*, or *Neorhizobium* and *Pseudorhizobium* were more closely related than the genus *Allorhizobium* and "*R. aggregatum* complex". Similarly, gANI and POCP values further supported the divergence of the members of *Allorhizobium* genus and "*R. aggregatum* complex" (Table S4). In this respect, members of the genus *Allorhizobium* exhibited gANI and POCP ranging 73.55-76.86 and 55.27-66.17, respectively, when compared with members of "*R. aggregatum* complex". In fact, the representatives of genera *Agrobacterium* were similarly closely related to *Neorhizobium* strains included into analysis (gANI 74.66-77.45; POCP 59.96-65.58).

Pan-genome analysis of *A. vitis* complex

Genome composition analyses of 14 *All. vitis* complex strains using GET_HOMOLOGUES software package resulted in 10,501 pan-genome clusters. The core genome (strict core and soft core compartments) of strains studied comprised 3,775 gene clusters (35.95% of total gene clusters), with 3,548 gene clusters present in all 14 strains (Fig. 3). The accessory genome contained 4,516 in the cloud (43% of total gene clusters) and

2,210 gene clusters in the shell (21.05% of total gene clusters) (Fig. 3). The accessory genes of particular clades of *All. vitis* complex were further analyzed as described below.

Clade-specific gene clusters

Gene clusters specific for particular clades of interests, comparing to related clades, were identified by either Pantagruel or GET_HOMOLOGUES software packages. Both sets of results were to a large extent congruent, although some differences were observed, owing to the distinct approaches employed by these software packages (data not shown). We focused on gene clusters for which we could predict putative function. The results are summarized below and in Table S5.

All. vitis species complex

There are 206 genes specific to the *All. vitis* complex (present in all *All. vitis sensu stricto*, *All. ampelinum* and *Allorhizobium* sp. Av2, and in no other *Allorhizobium*), which are mostly located on the second chromosome; while some specific genes are reported on the Ti plasmid and include the type 4 secretion system, this likely only reflect a sampling bias whereby all *All. vitis* complex isolates possess Ti plasmid. However, genes directly associated with pathogenicity, as well as Ti plasmid-associated genes and traits were not a subject of this study.

Half of the *All. vitis* complex-specific genes are gathered in clusters (Table S5), the rest are mostly scattered on chromosome 1 and have unknown function. Analysis of the predicted functions of clustered genes reveal a striking convergence of their function: most are involved in either environmental signal perception (4 clusters), stress response (2 clusters), aromatic compound and secondary metabolite biosynthesis (3 clusters) and/or aromatic compound degradation response (2 clusters). In addition, one cluster encodes a multicomponent K^+:H^+ antiporter, which is likely useful for adaptation to pH changes, and three clusters harbor several ABC transporter systems for sugar and nucleotide uptake. Finally, one cluster on the chromosome 1 encode a putative auto-transporter adhesin protein, which may have a role in commensalism and pathogenesis.

All *All. vitis* species complex strains studied carried a *pehA* gene encoding enzyme polygalacturonase. Unlike other agrobacteria, *All. vitis* strains are able to produce enzyme polygalacturonase, regardless of their tumorigenicity (McGuire et al., 1991). However, the presence of this gene has been determined also in *All. taibaishanense* 14971^T, *All. terrae* CC-HIH110^T and *All. oryziradicis* N19^T, but not in *All. undicola* ORS 992^T. Moreover, our bioinformatics analysis suggested the absence of this gene in other *Rhizobiaceae* members analyzed.

Furthermore, we detected the presence of gene encoding enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase (*acdS*) in all *All. vitis* species complex strains studied. This gene is considered to be important for plant-bacteria interaction through its involvement in lowering the level of ethylene produced by the plant (Gamalero &

Glick, 2015). This gene was found in other *Allorhizobium* spp., and some other Rhizobiaceae members analyzed in this study (data not shown), including *R. rhizogenes* strains. However, *acdS* gene was not present in *Agrobacterium* spp., even when the similarity search (blastp) was extended to *Agrobacterium* spp. strains available in GenBank, consistent with previous reporting (Bruto et al., 2014).

Tartrate utilization ability was reported for most of the *All. vitis* strains (Salomone et al., 1998, Szegedi, 1985, Salomone et al., 1996). Therefore, we mined *All. vitis* species complex genomes for the presence of tartrate utilization (TAR) regions. All strains except IPV-BO 6186 and IPV-BO 7105 carried tartrate utilization gene clusters. Moreover, we could not establish the presence of *All. vitis*-like TAR regions in any other Rhizobiaceae strains that we analyzed. Tartrate gene regions of *All. vitis* species complex strains were differentiated into four groups based on their sequence comparison analysis conducted by ANIb algorithm (Fig. S5; Table S6). The first group contained previously characterized TAR region called TAR-I, carried on tartrate utilization plasmid pTrAB3 of strain AB3 (Salomone et al., 1996, Szegedi et al., 1992). The second group carried representatives of regions TAR-II (carried on pTiAB3) and TAR-III (carried on pTrAB4) which are related to each other (Salomone et al., 1996, Crouzet & Otten, 1995). Another group was characterized by the TAR region lacking a second copy of *ttuC* gene (tartrate dehydrogenase), which we designated here as TAR-IV. This fourth group included the strain S4, in which the tartrate utilization system is located on a large plasmid pAtS4c (initially named pTrS4) (Szegedi et al., 1992). The TAR system of the strain Av2 was related to group TAR-I, but was characterized by the absence of *ttuA* gene (LysR-like regulator). We also compared strains studied by the distribution of particular TAR regions in their genomes (Table S7). *All. vitis sensu stricto* strains K309^T and KFB 253 carry TAR-II/III region. In addition to TAR-II/III region, strain KFB 239 contains TAR-1 region (Table S7), as well-characterized strain AB3 (Salomone et al., 1996). *All. ampelinum* strains S4^T, IPV-BO 1861-5, KFB 264 and V80/94 contain TAR-IV region, while the remaining strains IPV-BO 5159, KFB 243, KFB 250 and KFB 254 additionally carry TAR-II/III region (Table S7).

All. vitis sensu stricto

All five *All. vitis sensu stricto* strains analyzed in this study, but none of the *All. ampelinum* strains carried a gene cluster (Av-GC1, Table S5) with genes functionally annotated to be involved in the degradation process of salicylic acid and gentisic acid (2,5-dihydroxybenzoic acid) (MetaCyc pathways PWY-6640 and PWY-6223). The gene cluster Av-GC1 was located on Contig 1 (LMVL02000001.1) of reference strain K309^T, which seem to be part of the chromid, considering its high ANI with the chromid (Chromosome 2) of strain S4, whose genome sequence is finished. This gene cluster was also conserved in all other *All. vitis sensu stricto* strains available in GenBank (sub-clade A in Fig. S1), but was not identified in other *All. vitis* complex strains (clades B, C, D and D in Fig. S1). BLAST searches showed that this gene cluster is also present in some representatives of *Agrobacterium deltaense*, i.e. *Agrobacterium* genomospecies G7 (data not shown). In particular, genes comprising the Av-GC1 cluster encode the degradation of salicyl-CoA, an intermediate in degradation of

salicylic acid, to 3-fumarylpyruvate, via gentisic acid. Interestingly, strains KFB 239, IPV-BO 6186 and IPV-BO 7105 carried additional genes encoding the degradation of salicylaldehyde to salicyl-CoA via salicylic acid and salicyl adenylate, as well as the gene encoding the final step of gentisic acid degradation, the conversion of 3-fumarylpyruvate to fumarate and pyruvate. The three strains encoding enzymes of the complete pathway for degradation of salicylic acid and gentisic acid, and remaining strains K309^T and KFB 253 carrying a partial gene cluster, were phylogenetically separated and formed distinct sub-clades of *All.* vitis *sensu stricto* (Fig. 1).

All studied *All. vitis sensu stricto* strains also carried a gene cluster (Av-GC4, Table S5) annotated to be involved in the degradation of 4-hydroxyphenylacetate (MetaCyc pathway 3-HYDROXYPHENYLACETATE-DEGRADATION-PWY), which was not identified in any of the studied strains from sister species *All. ampelinum*. Nevertheless, this gene cluster was present in another genomic species of *All. vitis* complex of sub-clade C, with the exception of strain CG1013 (Fig. S1). The incomplete gene cluster was also carried by strain CG957 (sub-clade E; Fig. S1) and *Allorhizobium oryziradicis* N19. Gene content and comparative analysis suggested that cluster Av-GC4 is carried on a putative plasmid of *All. vitis sensu stricto* (data not shown).

Furthermore, *All. vitis sensu stricto* strains carried gene clusters (Av-GC2 and Av-GC3, Table S5) that might be involved in amino-acid uptake, catabolism and regulation, and putrescine uptake and degradation. However, we were not able to predict their function and compound associated with these clusters more precisely. In any case, both gene clusters likely are located on a putative plasmid, considering the presence of plasmid-specific genes (replication and/or conjugation associated genes) on the same contigs.

All. ampelinum

Among the 97 genes specific to *All. ampelinum* (present in all strains and in none of *All. vitis sensu stricto*), 60 genes are arranged in 15 clusters. Taking advantage of the finished genome status of strain S4, we found that 52/97 specific genes (9/15 gene clusters) occur on plasmids rather than chromosomes. This is a significant over-representation compared to the distribution of all genes (21.4% on plasmids, Chi-squared test p-value < 10^{-6}) or core genome genes (5.8% on plasmids, Chi-squared test p-value < 10^{-16}). The specific gene clusters encode a variety of putative biological functions; analysis of enrichment of high-level functional annotations revealed the over-representation of genes involved in the transport and metabolism of amino-acids or polyamines like putrescine (three separate clusters), lysin biosynthesis (two separate clusters), and nickel assimilation. The latter function is predicted for gene cluster Aa-GC15, which is located on the 631 kb-long megaplasmid pAtS4e and encodes the NikABCDE Ni^{2+} import system, a nickel-responsive transcriptional regulator NikR, as well as a set of chaperones and thioredoxins that may be involved in the biosynthesis of ion-associated cofactors. We also identified a gene cluster (Aa-GC3; Table S5) comprising putative genes involved in agrocinopine catabolism, a function we also identified

in other putative genes located next to this gene cluster. Interestingly, this gene cluster was located on chromosome 1 of reference strain S4.

Phenotypic and MALDI-TOF MS characterization

The phenotypic properties of the newly described species *All. ampelinum* are listed in Table 2. API 20NE and Biolog GEN III analyses did not reveal clear discriminative features between *All. vitis sensu stricto* and *All. ampelinum*. However, a weak positive reaction for 4-hydroxyphenylacetic (p-hydroxy-phenylacetic) acid for strains belonging to *All. vitis sensu stricto* was recorded, unlike for those belonging to *All. ampelinum* which were clearly negative. As bioinformatics analysis suggested that *All. vitis sensu stricto* strains carry a gene cluster encoding degradation of 4-hydroxyphenylacetate, the metabolism of this compound was assayed in a separate biochemical test. Our results indicated that all *All. vitis sensu stricto* strains tested are able to metabolize 4-hydroxyphenylacetate, which was recorded by a vigorous bacterial growth and a change of a medium color from turquoise (pH ~7.2) to yellow-green color (pH ~6.5), indicating the production of acid from the oxidation of substrate. On the other hand, *All. ampelinum* strains showed poor growth under culturing conditions, without change of medium color.

Although it was determined that *All. vitis sensu stricto* strains carry genes annotated to be involved in a degradation process of gentisic acid, this property could not be demonstrated in this study. Gentisic acid degradation genes could have lost their function or not be induced under our test conditions. Alternatively, the predicted function might be incorrect and the target substrate of these enzymes may be an unidentified compound more or less closely related to gentisic acid.

We also tested the ability of *All. vitis* complex strains to metabolize L-tartaric acid and produce alkali from this compound. In the present study, we included only strains that were not tested in our former work (Kuzmanović et al., 2014). Taken together, all *All. vitis* complex strains tested (Table 1) were able to produce alkali from tartrate and to change the color of the medium from turquoise to blue. Interestingly, strains IPV-BO 6186 and IPV-BO 7105, for which we could not identify tartrate utilization gene clusters, were also positive in this test.

As a broader way to characterize and distinguish the phenotype of strains we used MALDI-TOF mass-spectrometry (MS) of pure bacterial cultures. MALDI-TOF MS revealed diversity among strains studied (Fig. S4). Congruently with genomic data, this method allowed to separate strains from *All. vitis sensu stricto*, *All. ampelinum* and strain Av2, but also to distinguish strains within species (Fig. S4).

DISCUSSION

***Allorhizobium vitis* is not a single species**

Whole-genome sequencing and genomics allowed us to unravel the substantial taxonomic diversity within *All. vitis*. In particular, whole-genome sequence comparisons and phylogenomic analysis clearly showed that *Allorhizobium vitis* is not a single species, but

represents a species complex composed of several genomic species. Similarly, *Agrobacterium* biovar 1 was also initially considered a single species, but was later designated as a species complex comprising divergent genomic species. Several studies applying DDH initially demonstrated species diversity within *Agrobacterium* biovar 1 (De Ley, 1974, De Ley et al., 1973, Popoff et al., 1984), which was later supported by results obtained with AFLP (Portier et al., 2006, Mougel et al., 2002), housekeeping gene analysis (Costechareyre et al., 2010, Mousavi et al., 2014, Mousavi et al., 2015) and whole genome analysis (Lassalle et al., 2017). Although Ophel and Kerr (1990) also performed DDH for several *All. vitis* strains, diversity within this species remained unknown because the latter authors apparently analyzed closely related strains corresponding to *All. vitis sensu stricto* defined here.

Our previous study based on the analysis of several housekeeping gene sequences suggested the existence of several phylogenetic groups within *All. vitis* species complex (Kuzmanović et al., 2015). The present study focused on two phylogenetic groups defined in the previous study, the first comprises the type strain of *All. vitis* (strain K309^T) (Ophel & Kerr, 1990, Gan et al., 2018), whereas the second includes the well-characterized and completely sequenced strain S4 (Slater et al., 2009). Consequently, we amended the description of *All. vitis*, which now refers to the limited group of *All. vitis* complex strains (*All. vitis sensu stricto*) and proposed a description of a novel species *All. ampelinum* sp. nov. (see below formal description).

As indicated by the genome analysis of a larger set of strains available from NCBI GenBank database, the taxonomic diversity of *All. vitis* species complex is not limited to *All. vitis sensu stricto* and *All. ampelinum* sp. nov. However, the description of other sub-clades as a separate species was not performed in this study, because the sequencing of other strains was not conducted by our group and their draft genome sequences are still unpublished. Additionally, it is not clear if the sub-clades C and D (Fig. S1) represent a single or separate species. Further studies and isolation of diverse members of these clades are required to elucidate relationships between sub-clades C and D (Fig. S1). Overall, we kept the taxonomic status of sub-clades C, D and E (Fig. S1) as pending.

Specific functions and ecologies suggested by clade-specific gene cluster analysis

The convergence of functions encoded by the *All. vitis* species complex-specific genes suggests an ancient adaptation to different kind of stresses, including exposure to aromatic compounds, competition with other rhizospheric bacteria and pH change. The occurrence of multiple signal perception systems in the *All. vitis* species complex-specific indicate that adaptation to a changing environment seems a key feature of their ecology.

Moreover, we also mined genomes of *All. vitis* species complex strains for genes and gene clusters that were previously reported as important for the ecology of this bacterium. In this regard, polygalacturonase production, a trait associated with grapevine root necrosis (McGuire et al., 1991, Rodriguez-Palenzuela et al., 1991, Brisset et al., 1991) and tartrate degradation (Salomone et al., 1998) were proposed to contribute to the specialization of *All. vitis* to its grapevine host. Additionally, polygalacturonase activity might be involved in the process of the invasion of the host plant, as postulated before for some rhizobia (Muñoz et al., 1998). Although all *All. vitis* species complex strains carried *pehA* gene encoding

polygalacturonase, this gene was not restricted to this bacterial group, but was also present in other *Allorhizobium* spp. included in our analysis, except for *All. undicola*.

All *All. vitis* species complex strains included in this study, except for strains IPV-BO 6186 and IPV-BO 7105, carried TAR regions. However, all of them were able to metabolize tartrate and produce alkali from this compound. Therefore, strains IPV-BO 6186 and IPV-BO 7105 might carry another type of tartrate utilization system, distinct from those described so far in other *All. vitis* strains. Furthermore, diversity between TAR regions and variable distribution patterns of different TAR regions among strains were observed, in line with data reported previously (Salomone et al., 1996). The existence of non-tartrate-utilizing strains was also documented in the literature (Salomone et al., 1996). Considering the fact that the tartrate utilization in *All. vitis* is plasmid-borne (Szegedi et al., 1992, Otten et al., 1995, Crouzet & Otten, 1995), the acquisition of plasmid encoding this trait may provided the host strain with the selective advantage in tartrate-abundant ecological niches. Indeed, because grapevine is rich in tartrate (Ruffner, 2016), utilization of this substrate may enhance the competitiveness of *All. vitis* complex strains in colonizing this plant species (Salomone et al., 1998).

Furthermore, we observed that an important fraction of the species-specific genes for *All. vitis sensu stricto* and *All. ampelinum* occurred on chromids and plasmids, suggesting that these replicons may be an important part of the species' adaptive core genome, as previously observed in *A. tumefaciens* species complex (Lassalle et al., 2017). Ecological differentiation of the two main species of the complex seems to rely on consumption of different nutrient sources, including polyamines and nickel ion (potentially as a key cofactor of ecologically important enzymes) for *All. ampelinum*, and phenolic compounds for *All. vitis sensu stricto*.

Even though *All. vitis sensu stricto* strains carried a putative gene cluster which predicted function was the degradation of gentisic acid, we could not experimentally demonstrate this trait. Gentisic acid was detected in grapevine leaves (Pantelić et al., 2017) and is likely present in other parts of this plant. This compound was reported as a plant defense signal that can accumulate in some vegetable plants responding to compatible viral pathogens (Bellés et al., 1999, Bellés et al., 2006). In addition, a sub-clade within *All. vitis sensu stricto* composed of strains K309^{T} and KFB 253 carried a complete pathway for degradation of salicylic acid through gentisic acid. Salicylic acid is recognized as an important molecule for plant defense against certain pathogens (Vlot et al., 2009). The role of salicylic and gentisic acid in grapevine defense mechanism against pathogenic bacteria has not been studied in detail, and further investigations are required to understand their effect against tumorigenic agrobacteria. Furthermore, we demonstrated that all studied *All. vitis sensu stricto* strains can degrade 4-hydroxyphenylacetate, an activity that may contribute to the detoxication of aromatic compound and thus the survival of this bacterium in soil.

Similarly, gene clusters putatively involved in polyamine metabolism or nickel assimilation might confer to *All. ampelinum* the ability to persist in harsh environments. In this respect, nickel import has been shown to be essential for hydrogenase function in *E. coli* (Rowe et al., 2005). Hydrogenase function has in turn been proposed as a potential mechanism for detoxication of phenolic compounds in *A. vitis* (Biggs, 1994) and may thus have an important role in survival in the rhizosphere.

Surprisingly, we identified putative genes associated with catabolism of the opine agrocinopine harbored on a large circular chromosome of the reference strain S4. Thus far, such genes were reported as a plasmid-borne (Clare et al., 1990, Wetzel et al., 2014, Suzuki et al., 2000, Shao et al., 2018, Zhu et al., 2000, Kuzmanović & Puławska, 2019, Shao et al., 2019). Therefore, their presence on a chromosome might be a result of the horizontal gene transfer.

Status of the genus *Allorhizobium* and delineation of genera within the family *Rhizobiaceae*

On one hand, the genus *Allorhizobium* was clearly differentiated from other *Rhizobiaceae* genera on the basis of core- and pan-genome phylogenies (Figs. 1, 2, S1, S2 and S3), which was in line with previous studies employing genome-wide phylogeny (Ormeño-Orrillo et al., 2015, Hördt et al., 2020). We included diverse *All. vitis* complex strains into the analysis, showing that these bacteria, principally recognized as grapevine crown gall causative agents, belong to the genus *Allorhizobium*.

On the other hand, the taxonomic status of the "*R. aggregatum* complex" is still unresolved. Although MLSA suggested that this clade is a sister clade of the genus *Agrobacterium*, more thorough phylogenetic analysis performed in this study actually showed that the "*R. aggregatum* complex" is more closely related to the genus *Allorhizobium* (Figs. 1, 2, S1, S2 and S3). Presently, there is no widely accepted criteria and scientific consensus regarding the delineation of new bacterial genera (de Lajudie et al., 2019). In this study, existing *Rhizobiaceae* genera were compared using several demarcation methods proposed in the literature, such as AAI (Konstantinidis et al., 2017, Konstantinidis & Tiedje, 2007), POCP (Qin et al., 2014) and gANI/AF (Barco et al., 2020), which we complemented with genome-based phylogenies. Taken together, our genome-based analysis suggested that *Allorhizobium* and "*R. aggregatum* complex" should be considered as separate genera of the family *Rhizobiaceae*, and that "*R. aggregatum* complex" should represent a novel genus. Considering the grouping of the *Ciceribacter* spp. within this clade, a convenient solution might be an emendation of the genus *Ciceribacter* that would include members of the "*R. aggregatum* complex". However, a separate and more focused analysis of this clade is required to settle on this issue.

Emended description of *Allorhizobium vitis* (Ophel and Kerr 1990) Mousavi et al. 2016 emend. Hördt et al. 2020

The description of *Agrobacterium vitis* is provided by Ophel and Kerr (1990). Young et al. (2001) proposed the transfer of *A. vitis* to the genus *Rhizobium*, but it was neither widely accepted by a scientific community nor supported by further studies (Farrand et al., 2003, Lindström & Young, 2011, Costechareyre et al., 2010). Mousavi et al. (2015) reclassified this species to the genus *Allorhizobium*, which was included into the Validation list no. 172 of the IJSEM (Oren & Garrity, 2016). Hördt et al. (2020) emended a description of *All. vitis* by including genome sequence data for a type strain, which was published in the List of changes in taxonomic opinion no. 32 (Oren & Garrity, 2020).

As shown in this study, *All. vitis sensu stricto* includes a limited group of strains that can be differentiated from other *All. vitis* genomic species and other *Allorhizobium* species based on OGRIs, such as ANI, as well as by core-genome phylogeny. Moreover, *All. vitis sensu stricto* can be differentiated from other *All. vitis* complex strains by analysis of sequences of housekeeping genes *dnaK*, *gyrB* and *recA* (Kuzmanović et al., 2015). Moreover, this study demonstrated that strains belonging to this species can be distinguished from *All. ampelinum* by MALDI-TOF MS analysis. Unlike *All. ampelinum*, *All. vitis sensu stricto* strains are able to produce acid in a medium containing 4-hydroxyphenylacetate. However, this is considered as an accessory trait borne by a plasmid.

The whole-genome sequence of type strain K309^T is available in GenBank under the accession number LMVL00000000.2 and GCA_001541345.2 for the Nucleotide and Assembly databases, respectively (Gan et al., 2018). The genomic G+C content of the type strain is 57.55%. Its approximate genome size is 5.75 Mbp.

Basonym: *Agrobacterium vitis* Ophel and Kerr 1990.

The type strain, K309^T (= NCPPB 3554^T =HAMBI 1817^T = ATCC 49767^T = CIP 105853^T = ICMP 10752^T = IFO 15140^T = JCM 21033^T = LMG 8750^T = NBRC 15140^T). was isolated from grapevine in South Australia in 1977.

Description of *Allorhizobium ampelinum* sp. nov.

The description and properties of the new species are given in Table 2.

All. ampelinum (am.pe.li'num. Gr. n. ampelos grapevine; Gr. adj. ampelinos and N.L. neut. adj. ampelinum of the vine).

All. ampelinum strains were formerly classified to belong to the species *All. vitis*. However, our data indicated that they can be distinguished from *All. vitis sensu stricto* and other *All. vitis* genomic species based on OGRIs (e.g. ANI and isDDH) and core-genome phylogeny, as well as by analysis of sequences of housekeeping genes (Kuzmanović et al., 2015). Furthermore, *All. ampelinum* can be differentiated from *All. vitis sensu stricto* by MALDI-TOF MS analysis.

The type strain, S4^T (= DSM 112012^T = ATCC BAA-846^T) was isolated from grapevine tumor in Hungary in 1981.

ACKNOWLEDGMENTS

This research was supported by the Georg Forster Fellowship for postdoctoral research from the Alexander von Humboldt-Foundation, Bonn, Germany. Bioinformatic analyses were supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B). We would like to thank Cathrin Spröer and Boyke Bunk for conducting the Illumina sequencing. We are grateful to Anja Frühling and Ulrike Steiner for support in phenotypic tests.

AUTHOR CONTRIBUTIONS

N.K. and F.L. conceived and designed the study, and analyzed data. S.V. performed phenotypic characterization of the strains (API, Biolog and MALDI-TOF). E.B. provided bacterial strains. J.P., J.O. and K.S were involved in interpreting data. N.K. and F.L. wrote the manuscript. All authors read, discussed, edited and approved the final manuscript.

REFERENCES

Afgan E, Baker D, Batut B*, et al.*, 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* **46**, W537-W44.

Alexa A, Rahnenführer J, Lengauer T, 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-7.

Bankevich A, Nurk S, Antipov D*, et al.*, 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455-77.

Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson D, 2020. A genus definition for *Bacteria* and *Archaea* based on a standard genome relatedness index. *mBio* **11**, e02475-19.

Bellés JM, Garro R, Fayos J, Navarro P, Primo J, Conejero V, 1999. Gentisic acid as a pathogen-inducible signal, additional to salicylic acid for activation of plant defenses in tomato. *Molecular Plant-Microbe Interactions®* **12**, 227-35.

Bellés JM, Garro R, Pallás V, Fayos J, Rodrigo I, Conejero V, 2006. Accumulation of gentisic acid as associated with systemic infections but not with the hypersensitive response in plant-pathogen interactions. *Planta* **223**, 500-11.

Biggs J, 1994. *Ecology and biological control of Agrobacterium vitis, the grapevine crown gall pathogen*. Adelaide: University of Adelaide, PhD Thesis.

Bini F, Geider K, Bazzi C, 2008a. Detection of *Agrobacterium vitis* by PCR using novel *virD2* gene-specific primers that discriminate two subgroups. *European Journal of Plant Pathology* **122**, 403-11.

Bini F, Kuczmog A, Putnoky P*, et al.*, 2008b. Novel pathogen-specific primers for the detection of *Agrobacterium vitis* and *Agrobacterium tumefaciens*. *Vitis* **47**, 181-9.

Bishop AL, Burr TJ, Mittak VL, Katz BH, 1989. A monoclonal antibody specific to *Agrobacterium tumefaciens* biovar 3 and its utilization for indexing grapevine propagation material. *Phytopathology* **79**, 995-8.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20.

Brisset M-N, Rodriguez-Palenzuela P, Burr TJ, Collmer A, 1991. Attachment, chemotaxis, and multiplication of *Agrobacterium tumefaciens* biovar 1 and biovar 3 on grapevine and pea. *Applied and Environmental Microbiology* **57**, 3178-82.

Bruto M, Prigent-Combaret C, Muller D, Moënne-Loccoz Y, 2014. Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria. *Scientific Reports* **4**, 6261.

Caspi R, Altman T, Billington R*, et al.*, 2013. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research* **42**, D459-D71.

Chun J, Rainey FA, 2014. Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. *International Journal of Systematic and Evolutionary Microbiology* **64**, 316-24.

Clare BG, Kerr A, Jones DA, 1990. Characteristics of the nopaline catabolic plasmid in *Agrobacterium* strains K84 and K1026 used for biological control of crown gall disease. *Plasmid* **23**, 126-37.

Contreras-Moreira B, Vinuesa P, 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology* **79**, 7696-701.

Costechareyre D, Rhouma A, Lavire C*, et al.*, 2010. Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis: *Agrobacterium recA* diversity. *Microbial Ecology* **60**, 862-72.

Crouzet P, Otten L, 1995. Sequence and mutational analysis of a tartrate utilization operon from Agrobacterium vitis. *Journal of Bacteriology* **177**, 6518-26.

De Lajudie P, Laurent-Fulele E, Willems A*, et al.*, 1998. *Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal. *International Journal of Systematic and Evolutionary Microbiology* **48**, 1277-90.

De Lajudie PM, Andrews M, Ardley J*, et al.*, 2019. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *International Journal of Systematic and Evolutionary Microbiology* **69**, 1852-63.

De Ley J, 1974. Phylogeny of procaryotes. *Taxon* **23**, 291–300.

De Ley J, Tijtgat R, De Smedt J, Michiels M, 1973. Thermal stability of DNA: DNA hybrids within the genus *Agrobacterium*. *Journal of General Microbiology* **78**, 241-52.

El-Gebali S, Mistry J, Bateman A*, et al.*, 2018. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427-D32.

Farrand SK, Van Berkum PB, Oger P, 2003. *Agrobacterium* is a definable genus of the family *Rhizobiaceae*. *International Journal of Systematic and Evolutionary Microbiology* **53**, 1681-7.

Fuller SL, Savory EA, Weisberg AJ*, et al.*, 2017. Isothermal amplification and lateral-flow assay for detecting crown-gall-causing *Agrobacterium* spp. *Phytopathology* **107**, 1062-8.

Gamalero E, Glick BR, 2015. Bacterial modulation of plant ethylene levels. *Plant Physiology* **169**, 13-22.

Gan HM, Lee MVJ, Savka MA, 2018. High-quality draft genome sequence of the type strain of *Allorhizobium vitis*, the primary causal agent of grapevine crown gall. *Microbiology resource announcements* **7**, e01045-18.

Goris J, Konstantinidis K, Klappenbach J, Coenye T, Vandamme P, Tiedje J, 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**.

Harrison PW, Lower RP, Kim NK, Young JP, 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology* **18**, 141-8.

Hördt A, López MG, Meier-Kolthoff JP*, et al.*, 2020. Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria. *Frontiers in Microbiology* **11**.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S, 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**, 5114.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M, 2016a. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457-D62.

Kanehisa M, Sato Y, Morishima K, 2016b. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology* **428**, 726-31.

Kerr A, Panagopoulos CG, 1977. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. *Journal of Phytopathology* **90**, 172-9.

Konstantinidis KT, Rossello-Mora R, Amann R, 2017. Uncultivated microbes in need of their own taxonomy. *Isme j* **11**, 2399-406.

Konstantinidis KT, Tiedje JM, 2005. Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology* **187**, 6258-64.

Konstantinidis KT, Tiedje JM, 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology* **10**, 504-9.

Koonin EV, Wolf YI, 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**, 6688-719.

Kuzmanović N, Biondi E, Bertaccini A, Obradović A, 2015. Genetic relatedness and recombination analysis of *Allorhizobium vitis* strains associated with grapevine crown gall outbreaks in Europe. *Journal of Applied Microbiology* **119**, 786-96.

Kuzmanović N, Ivanović M, Prokić A, Gašić K, Zlatković N, Obradović A, 2014. Characterization and phylogenetic diversity of *Agrobacterium vitis* from Serbia based on sequence analysis of 16S-23S rRNA internal transcribed spacer (ITS) region. *European Journal of Plant Pathology* **140**, 757-68.

Kuzmanović N, Puławska J, 2019. Evolutionary relatedness and classification of tumor-inducing and opine-catabolic plasmids in three *Rhizobium rhizogenes* strains isolated from the same crown gall tumor. *Genome Biology and Evolution* **11**, 1525-40.

Kuzmanović N, Pulawska J, Hao L, Burr TJ, 2018. The ecology of *Agrobacterium vitis* and management of crown gall disease in vineyards. *Current Topics in Microbiology and Immunology*.

Lassalle F, Dastgheib SMM, Zhao F-J*, et al.*, 2020. Phylogenomics reveals the basis of adaptation of Pseudorhizobium species to extreme environments and supports a taxonomic revision of the genus. *Systematic and Applied Microbiology*, In Press.

Lassalle F, Planel R, Penel S*, et al.*, 2017. Ancestral genome estimation reveals the history of ecological diversification in *Agrobacterium*. *Genome Biology and Evolution* **9**, 3413-31.

Lassalle F, Veber P, Jauneikaite E, Didelot X, 2019. Automated reconstruction of all gene histories in large bacterial pangenome datasets and search for co-evolved gene modules with Pantagruel. *bioRxiv*, 586495.

Li L, Stoeckert CJ, Jr., Roos DS, 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-89.

Lindström K, Young JPW, 2011. International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*: Minutes of the meeting, 7 September 2010, Geneva, Switzerland. *International Journal of Systematic and Evolutionary Microbiology* **61**, 3089-93.

Mcguire RG, Rodriguez-Palenzuela P, Collmer A, Burr TJ, 1991. Polygalacturonase production by *Agrobacterium tumefaciens* biovar 3. *Applied and Environmental Microbiology* **57**, 660-4.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M, 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 1-14.

Morton ER, Fuqua C, 2012. Laboratory maintenance of *Agrobacterium*. *Current Protocols in Microbiology* **24**, 3D.1.-3D.1.6.

Mougel C, Thioulouse J, Perriere G, Nesme X, 2002. A mathematical method for determining genome divergence and species delineation using AFLP. *International Journal of Systematic and Evolutionary Microbiology* **52**, 573-86.

Mousavi SA, Osterman J, Wahlberg N*, et al.*, 2014. Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. *Systematic and Applied Microbiology* **37**, 208-15.

Mousavi SA, Willems A, Nesme X, De Lajudie P, Lindström K, 2015. Revised phylogeny of *Rhizobiaceae*: Proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations. *Systematic and Applied Microbiology* **38**, 84-90.

Muñoz JA, Coronado C, Pérez-Hormaeche J, Kondorosi A, Ratet P, Palomares AJ, 1998. MsPG3, a *Medicago sativa* polygalacturonase gene expressed during the alfalfa-*Rhizobium meliloti* interaction. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9687-92.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ, 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268-74.

Ophel K, Kerr A, 1990. *Agrobacterium vitis* sp. nov. for strains of *Agrobacterium* biovar 3 from grapevines. *International Journal of Systematic Bacteriology* **40**, 236-41.

Oren A, Garrity G, 2020. Notification of changes in taxonomic opinion previously published outside the IJSEM. *International Journal of Systematic and Evolutionary Microbiology* **70**, 4061-90.

Oren A, Garrity GM, 2016. List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology* **66**, 4299-305.

Ormeño-Orrillo E, Servín-Garcidueñas LE, Rogel MA*, et al.*, 2015. Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. *Systematic and Applied Microbiology* **38**, 287-91.

Otten L, Crouzet P, Salomone JY, De Ruffray P, Szegedi E, 1995. *Agrobacterium vitis* strain AB3 harbors two independent tartrate utilization systems, one of which is encoded by the Ti plasmid. *Molecular Plant-Microbe Interactions* **8**, 138-46.

Panagopoulos CG, Psallidas PG, 1973. Characteristics of Greek isolates of *Agrobacterium tumefaciens* (E.F. Smith & Townsend) Conn. *Journal of Applied Bacteriology* **36**, 233-40.

Panagopoulos CG, Psallidas PG, Alivizatos AS. Studies on biotype 3 of *Agrobacterium radiobacter* var. *tumefaciens*. *Proceedings of the 4th International conference on plant pathogenic bacteria*, *1978*. Angers, France, 221-8.

Pantelić MM, Zagorac DČD, Ćirić IŽ*, et al.*, 2017. Phenolic profiles, antioxidant activity and minerals in leaves of different grapevine varieties grown in Serbia. *Journal of Food Composition and Analysis* **62**, 76-83.

Parks DH, Chuvochina M, Waite DW*, et al.*, 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**, 996-1004.

Popoff MY, Kersters K, Kiredjian M, Miras I, Coynault C, 1984. Taxonomic position of *Agrobacterium* strains of hospital origin. *Ann Microbiol (Paris)* **135a**, 427-42.

Portier P, Fischer-Le Saux M, Mougel C*, et al.*, 2006. Identification of Genomic Species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. *Applied and Environmental Microbiology* **72**, 7123-31.

Qin Q-L, Xie B-B, Zhang X-Y*, et al.*, 2014. A proposed genus boundary for the prokaryotes based on genomic insights. *Journal of Bacteriology* **196**, 2210.

Ramírez-Bahena MH, Vial L, Lassalle F*, et al.*, 2014. Single acquisition of protelomerase gave rise to speciation of a large and diverse clade within the *Agrobacterium*/*Rhizobium* supercluster characterized by the presence of a linear chromid. *Molecular Phylogenetics and Evolution* **73**, 202-7.

Richter M, Rossello-Mora R, 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19126-31.

Rodriguez-Palenzuela P, Burr TJ, Collmer A, 1991. Polygalacturonase is a virulence factor in *Agrobacterium tumefaciens* biovar 3. *Journal of Bacteriology* **173**, 6547-52.

Rowe JL, Starnes GL, Chivers PT, 2005. Complex transcriptional control links NikABCDE-dependent nickel transport with hydrogenase expression in *Escherichia coli*. *Journal of Bacteriology* **187**, 6317.

Ruffner HP, 2016. Metabolism of tartaric and malic acids in *Vitis* : A review - Part A. *Vitis: Journal of Grapevine Research* **21**, 247-.

Salomone JY, Crouzet P, De Ruffray P, Otten L, 1996. Characterization and distribution of tartrate utilization genes in the grapevine pathogen *Agrobacterium vitis*. *Mol Plant Microbe Interact* **9**, 401-8.

Salomone JY, Szegedi E, Cobanov P, Otten L, 1998. Tartrate utilization genes promote growth of *Agrobacterium* spp. on grapevine. *Molecular Plant-Microbe Interactions* **11**, 836-8.

Sawada H, Ieki H, 1992. Crown gall of kiwi caused by *Agrobacterium tumefaciens* in Japan. *Plant Disease* **76**, 212.

Schumann P, Maier T, 2014. Chapter 13 - MALDI-TOF mass spectrometry applied to classification and identification of bacteria. In: Goodfellow M, Sutcliffe I, Chun J, eds. *Methods in Microbiology.* Academic Press, 275-306. (41.)

Seemann T, 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-9.

Shao S, Van Heusden GPH, Hooykaas PJJ, 2019. Complete sequence of succinamopine Ti-plasmid pTiEU6 reveals its evolutionary relatedness with nopaline-type Ti-plasmids. *Genome Biology and Evolution* **11**, 2480-91.

Shao S, Zhang X, Van Heusden GPH, Hooykaas PJJ, 2018. Complete sequence of the tumor-inducing plasmid pTiChry5 from the hypervirulent *Agrobacterium tumefaciens* strain Chry5. *Plasmid* **96-97**, 1-6.

Slater SC, Goldman BS, Goodner B*, et al.*, 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *Journal of Bacteriology* **191**, 2501-11.

Stackebrandt E, Goebel BM, 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* **44**.

Süle S, 1978. Biotypes of *Agrobacterium tumefaciens* in Hungary. *Journal of Applied Bacteriology* **44**, 207-13.

Suzuki K, Hattori Y, Uraji M*, et al.*, 2000. Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. *Gene* **242**, 331-6.

Szegedi E, 1985. Host range and specific L(+)tartrate utilization of biotype 3 of *Agrobacterium tumefaciens*. *Acta Phytopathologica Academiae Scientiarum Hungaricae* **20**, 17-22.

Szegedi E, Otten L, Czakó M, 1992. Diverse types of tartrate plasmids in *Agrobacterium tumefaciens* biotype III strains. *Molecular Plant-Microbe Interactions* **5**, 435-8.

Tatusova T, Dicuccio M, Badretdin A*, et al.*, 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **44**, 6614-24.

Tempé J, Petit A, Holsters M, Montagu M, Schell J, 1977. Thermosensitive step associated with transfer of the Ti plasmid during conjugation: Possible relation to transformation in crown gall. *Proc Natl Acad Sci U S A* **74**, 2848-9.

Tóth EM, Schumann P, Borsodi AK, Kéki Z, Kovács AL, Márialigeti K, 2008. *Wohlfahrtiimonas chitiniclastica* gen. nov., sp. nov., a new gammaproteobacterium isolated from *Wohlfahrtia magnifica* (Diptera: Sarcophagidae). *International Journal of Systematic and Evolutionary Microbiology* **58**, 976-81.

Varghese NJ, Mukherjee S, Ivanova N*, et al.*, 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Research* **43**, 6761-71.

Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B, 2018. GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Frontiers in Microbiology* **9**.

Vlot AC, Dempsey DA, Klessig DF, 2009. Salicylic acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology* **47**, 177-206.

Wayne LG, Brenner DJ, Colwell RR*, et al.*, 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology* **37**.

Wetzel ME, Kim KS, Miller M, Olsen GJ, Farrand SK, 2014. Quorum-dependent mannopine-inducible conjugative transfer of an *Agrobacterium* opine-catabolic plasmid. *Journal of Bacteriology* **196**, 1031-44.

Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J, 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281-6.

Young JM, Kerr A, Sawada H, 2005. Genus *Agrobacterium* Conn 1942, 359[AL]. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, eds. *Bergey's Manual of Systematics of Archaea and Bacteria.* New York: Springer-Verlag, 340-5. (The Proteobacteria; vol.)

Young JM, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H, 2001. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *International Journal of Systematic and Evolutionary Microbiology* **51**, 89-103.

Zhu J, Oger PM, Schrammeijer B, Hooykaas PJ, Farrand SK, Winans SC, 2000. The bases of crown gall tumorigenesis. *Journal of Bacteriology* **182**, 3885-95.

Tables:

Table 1 Strains of "*A. vitis*" clade analyzed in this study and their genome sequence features

Strain	Species	Geographic origin	Year of isolation	Reference	Genome sequencing	Contigs (N)	N50 (Kb)	Size (Mb)	GC Content (%)	Gene[1]	CDS[1]	Accession number
K309[T]	*All. vitis sensu stricto*	Australia	1977	(Ophel & Kerr, 1990)	(Gan et al., 2018)	22	999	5.75	57.55	5188	5136	LMVL00000000.2
IPV-BO 6186	*All. vitis sensu stricto*	Italy	2006	(Bini et al., 2008a)	This study	79	608	5.80	57.57	5250	5196	VOLK00000000.1
IPV-BO 7105	*All. vitis sensu stricto*	Italy	2007	(Kuzmanović et al., 2015)	This study	91	462	5.81	57.54	5214	5157	VOLJ00000000.1
KFB 239	*All. vitis sensu stricto*	Serbia	2010	(Kuzmanović et al., 2014)	This study	82	456	6.15	57.57	5545	5490	VOLI00000000.1
KFB 253	*All. vitis sensu stricto*	Serbia	2011	(Kuzmanović et al., 2014)	This study	70	401	5.81	57.56	5290	5235	VOLF00000000.1
S4[T]	*All. ampelinum*	Hungary	1981	(Szegedi, 1985)	(Slater et al., 2009)	CG[2]	CG[2]	6.32	57.47	5840	5770	CP000633.1-CP000639.1
IPV-BO 1861-5	*All. ampelinum*	Italy	1984	(Bini et al., 2008b)	This study	110	971	5.67	57.63	5125	5072	VOLM00000000.1
IPV-BO 5159	*All. ampelinum*	Italy	2003	(Bini et al., 2008b)	This study	160	269	6.47	57.56	5939	5883	VOLL00000000.1
KFB 243	*All. ampelinum*	Serbia	2011	(Kuzmanović et al., 2014)	This study	210	270	6.52	57.62	5963	5909	VOLH00000000.1
KFB 250	*All. ampelinum*	Serbia	2011	(Kuzmanović et al., 2014)	This study	135	356	6.47	57.60	5845	5790	VOLG00000000.1
KFB 254	*All. ampelinum*	Serbia	2011	(Kuzmanović et al., 2014)	This study	76	492	5.96	57.60	5433	5380	VOLE00000000.1
KFB 264	*All. ampelinum*	Serbia	2011	(Kuzmanović et al., 2014)	This study	75	714	5.92	57.52	5355	5299	VOLD00000000.1
V80/94	*All. ampelinum*	USA	Oregon	(Fuller et al., 2017)	(Fuller et al., 2017)	66	413	5.98	57.48	5467	5414	NBZE00000000.1
Av2	*Allorhizobium* sp.	Croatia	2006	(Kuzmanović et al., 2015)	This study	123	407	6.23	57.58	5713	5659	VOLN00000000.1

[1] Numbers based on Prokka annotations.
[2] CG, Complete genome.

Table 2. Protologue for *Allorhizobium ampelinum* sp. nov.

Species name	*Allorhizobium ampelinum*
Genus name	*Allorhizobium*
Specific epithet	*ampelinum*
Species status	sp. nov.
Species etymology	am.pe.li'num. Gr. n. ampelos grapevine; Gr. adj. ampelinos and N.L. neut. adj. ampelinum of the vine
Designation of the type strain	S4
Strain collection numbers	DSM 112012^T, ATCC BAA-846^T
16S rRNA gene accession number	U28505.1
Genome accession number	GCF_000016285.1
Genome status	Complete
Genome size	6,320,946
GC mol %	57.47
Country of origin	Hungary
Region of origin	Orgovány, Bács-Kiskun county
Date of isolation	1981
Source of isolation	Aerial gall on 2-year-old woody grapevine (*Vitis vinifera* cv. 'Izsaki Sarfeher')
Sampling date	1981
Number of strains in study	8
Source of isolation of non-type strains	Grapevine
Growth medium, incubation conditions used for standard cultivation	Yeast mannitol agar (YMA) at 28°C
Conditions of preservation	-80°C
Gram stain	Negative
Cell shape	Rod
Colony morphology	Colonies on YMA are white to cream coloured, circular, convex and glistening
Positive tests with BIOLOG	pH 6, D-Mannose, D-Galactose, 1% Sodium Lactate, Pectin, Rifamycin SV, Tetrazolium Blue, Potassium Tellurite
Negative tests with BIOLOG	pH5, N-Acetyl Neuraminic Acid, 4% NaCl, 8% NaCl, 3-Methyl Glucose, Inosine, Fusidic Acid, Troleandomycin, D-Serine, Minocycline, Guanidine HCl, Niaproof 4, p-Hydroxy-Phenylacetic Acid, Lithium Chloride, γ-Amino-Butryric Acid, α-Hydroxy-Butyric Acid, α-Keto-Butyric Acid, Propionic Acid, Sodium Butyrate
Positive tests with API	URE, ESC, PNG, GLU (assimilation), ARA, MNE, MAN, MLT, OX
Negative tests with API	NO3, TRP, GLU (fermentation), ADH, GEL, CAP, ADI, PAC
Variable tests with API	NAG, MAL, GNT, CIT
Commercial kits used	BIOLOG GEN3, API 20NE
Oxidase[1]	Positive
Positive tests[1]	Growth at 35°C, growth in nutrient broth supplemented with 2% NaCl, citrate utilization, production of acid from sucrose, production of alkali from tartrate
Negative tests[1]	Production of 3-ketolactose from lactose, acid-clearing on PDA with $CaCO_3$, production of reddish-brown pellicle at the surface of ferric ammonium citrate broth, motility at pH 7.0, acid from d-(+)-melezitose, acid production from 4-hydroxyphenylacetate
Known pathogenicity	Plant pathogenic

[1]These tests were performed for strains S4^T, KFB 243, KFB 250, KFB 254 and KFB 264 by (Kuzmanović et al., 2014), except for a test of acid production in a medium containing 4-hydroxyphenylacetate conducted in this study. For strains that were not included in our former study, test of production of alkali from tartrate was conducted in the present work.

Figure captions:

Fig. 1 Maximum-likelihood core-genome phylogeny of the genus *Allorhizobium* and other Rhizobiaceae members. The tree was estimated with IQ-TREE from the concatenated alignment of 344 top-ranked genes selected using GET_PHYLOMARKERS software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQTREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R6 model. "*Allorhizobium vitis*" clade is collapsed on the left tree and shown expanded on the right. The matrix represents the calculated ANIb values for the genomic sequences. The same tree, but without collapsing clades, is presented in the Figure S2.

Fig. 2 Maximum-likelihood pan-genome phylogeny of the genus *Allorhizobium* and other Rhizobiaceae members. The tree was estimated with IQ-TREE from the consensus (COGtriangles and OMCL clusters) pan-genome matrix containing 33,396 clusters obtained using GET_ HOMOLOGUES software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQTREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR2+FO+R5 model. The same tree, but without collapsing clades, is presented in the Figure S3.

Fig. 3 Genome composition analyses of 14 *All. vitis* complex strains studied. **a** Bar plot showing the absolute size frequencies of orthologous clusters as predicted by the COGtriangles and OMCL algorithms. **b** Pie chart showing the relative sizes (cluster numbers) contained in the core, soft-core, shell, and cloud genomes.

Fig. S1 Maximum-likelihood core-genome phylogeny of the genus *Allorhizobium* (including 48 strains of *All. vitis* complex strains whose sequences are available in the GenBank) and other Rhizobiaceae members (Dataset 1). The tree was estimated with IQ-TREE from the concatenated alignment of 302 top-ranked genes selected using GET_PHYLOMARKERS software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQTREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R7 model. The matrix represents the calculated ANIb values for the genomic sequences.

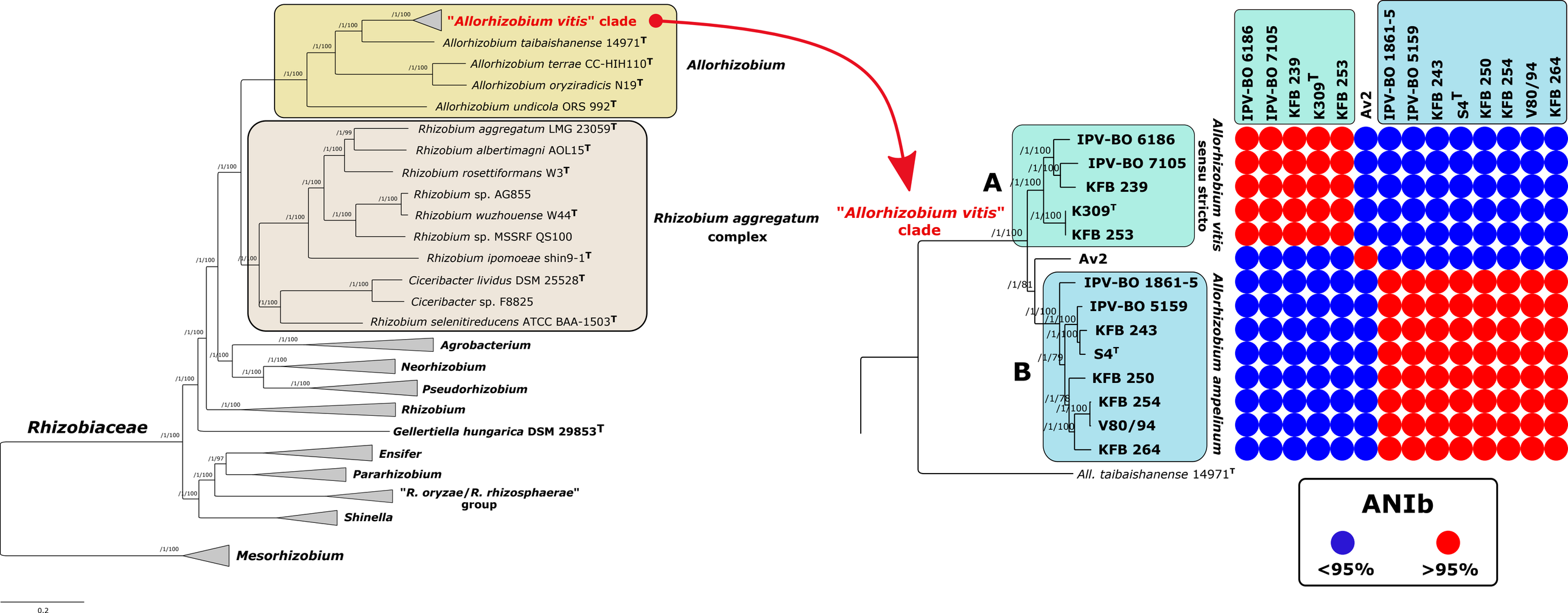
Fig. S2 Maximum-likelihood core-genome phylogeny of the genus *Allorhizobium* and other Rhizobiaceae members. The tree was estimated with IQ-TREE from the concatenated alignment of 344 top-ranked genes selected using GET_PHYLOMARKERS software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQTREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R6 model. The same tree, but with collapsing clades, is presented in the Figure 1.

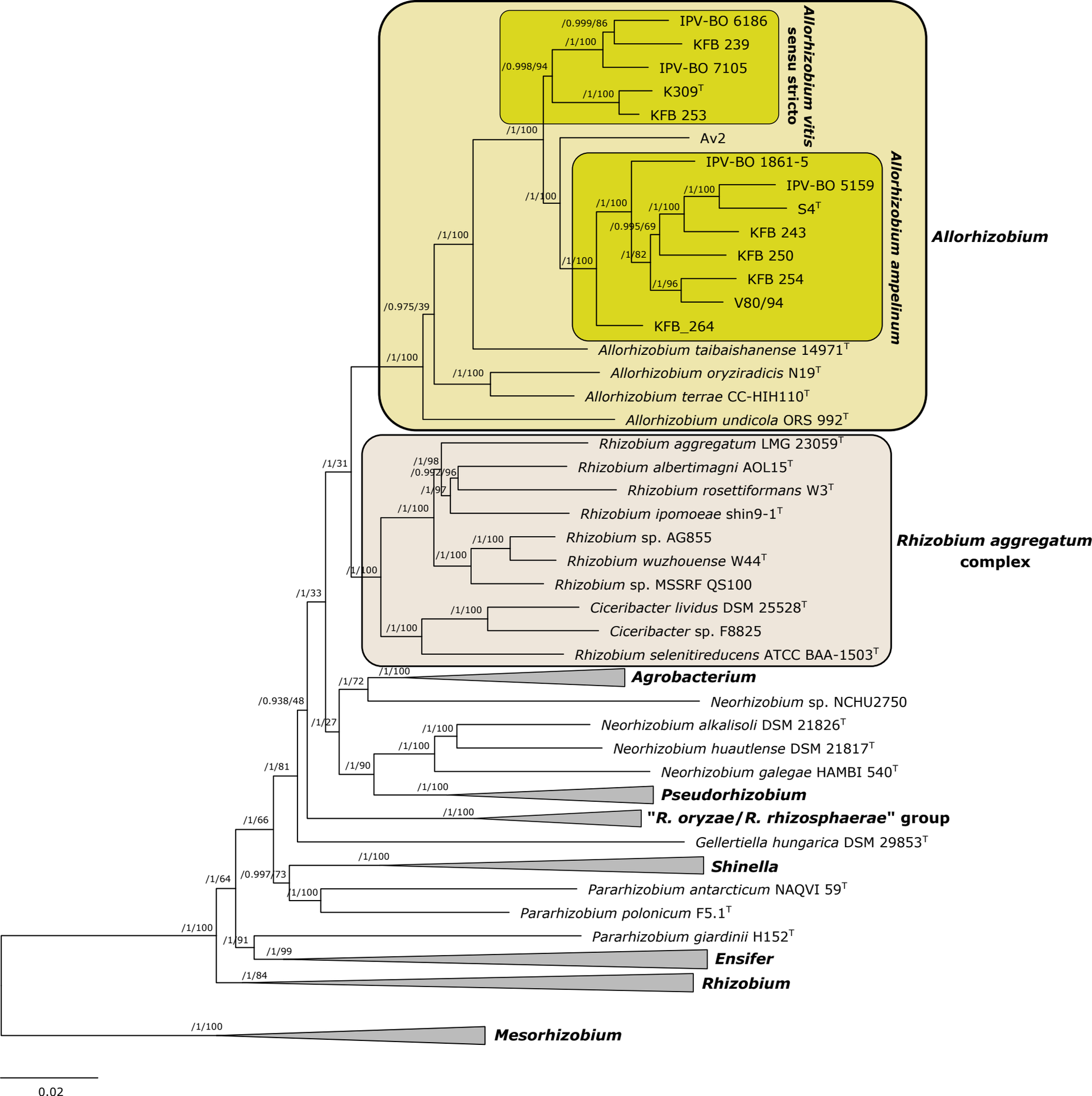
Fig. S3 Maximum-likelihood pan-genome phylogeny of the genus *Allorhizobium* and other Rhizobiaceae members. The tree was estimated with IQ-TREE from the consensus (COGtriangles and OMCL clusters) pan-genome matrix containing 33,396 clusters obtained using GET_ HOMOLOGUES

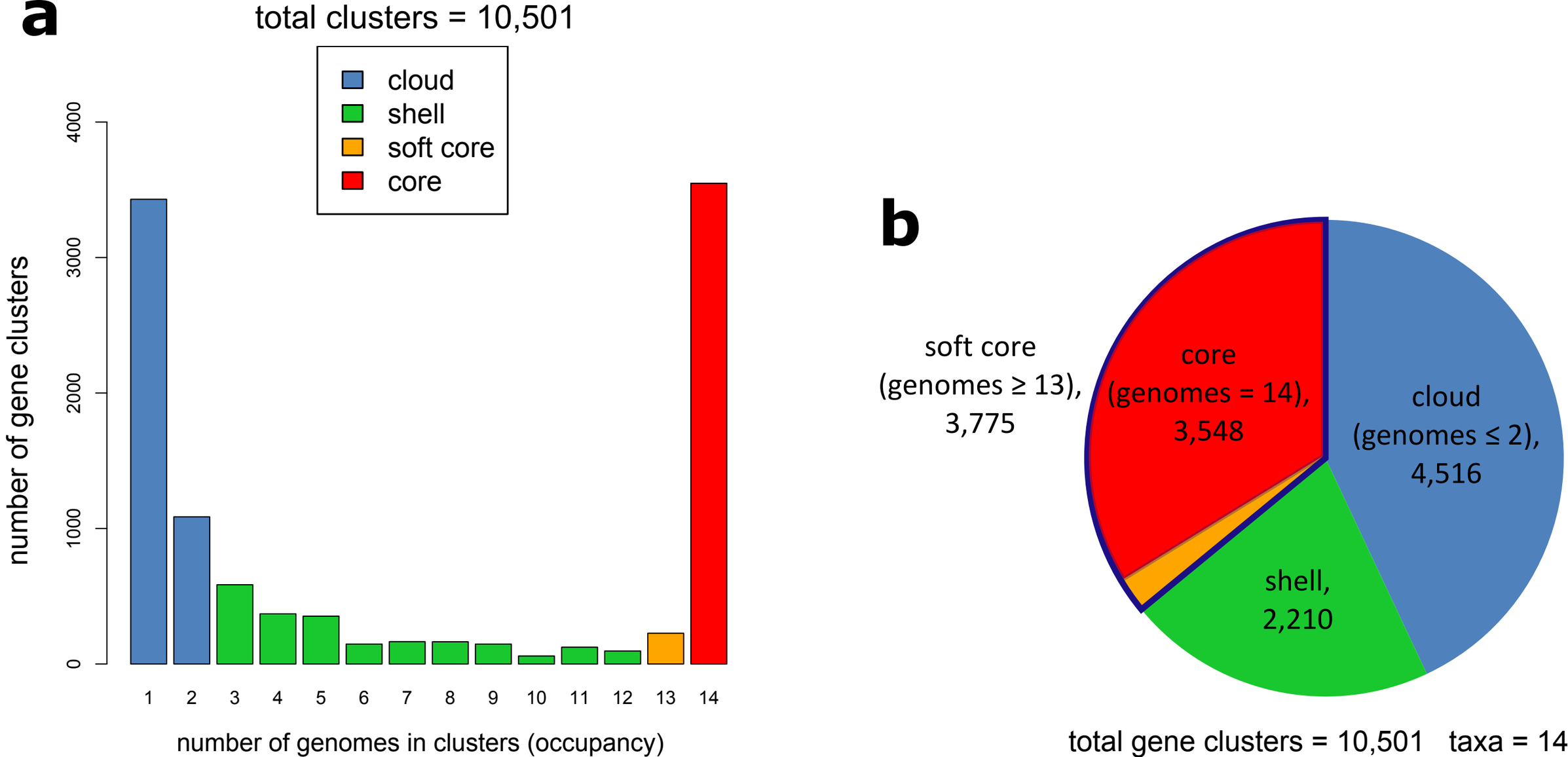
software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQTREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR2+FO+R5 model. The same tree, but with collapsing clades, is presented in the Figure 2.

Fig. S4 Score-oriented dendrogram showing the similarity of the MALDI-TOF mass spectra of 14 *All. vitis* complex strains studied. The dendrogram was created using the MALDI Biotyper Compass Explorer software (Bruker, Version 4.1.90).

Fig. S5 Heatmap representation of the average nucleotide identity (ANIb) for TAR regions of *All. vitis* complex strains. PyANI program Version 0.2.9 (https://github.com/widdowquinn/pyani) was used to calculate ANIb values and generate the clustered heatmap.


"Allorhizobium vitis" clade
Allorhizobium taibaishanense 14971T
Allorhizobium terrae CC-HIH110T
Allorhizobium oryziradicis N19T
Allorhizobium undicola ORS 992T
Allorhizobium
Rhizobium aggregatum LMG 23059T
Rhizobium albertimagni AOL15T
Rhizobium rosettiformans W3T
Rhizobium sp. AG855
Rhizobium wuzhouense W44T
Rhizobium sp. MSSRF QS100
Rhizobium ipomoeae shin9-1T
Ciceribacter lividus DSM 25528T
Ciceribacter sp. F8825
Rhizobium selenitireducens ATCC BAA-1503T
Rhizobium aggregatum complex
Agrobacterium
Neorhizobium
Pseudorhizobium
Rhizobium
Gellertiella hungarica DSM 29853T
Ensifer
Pararhizobium
"R. oryzae/R. rhizosphaerae" group
Shinella
Mesorhizobium
Rhizobiaceae
0.2
"Allorhizobium vitis" clade
A
B
IPV-BO 6186
IPV-BO 7105
KFB 239
K309T
KFB 253
Av2
IPV-BO 1861-5
IPV-BO 5159
KFB 243
S4T
KFB 250
KFB 254
V80/94
KFB 264
All. taibaishanense 14971T
Allorhizobium vitis sensu stricto
Allorhizobium ampelinum
ANIb
<95%
>95%


Allorhizobium vitis sensu stricto
IPV-BO 6186
KFB 239
IPV-BO 7105
K309T
KFB 253
Av2
Allorhizobium ampelinum
IPV-BO 1861-5
IPV-BO 5159
S4T
KFB 243
KFB 250
KFB 254
V80/94
KFB_264
Allorhizobium taibaishanense 14971T
Allorhizobium oryziradicis N19T
Allorhizobium terrae CC-HIH110T
Allorhizobium undicola ORS 992T
Allorhizobium
Rhizobium aggregatum LMG 23059T
Rhizobium albertimagni AOL15T
Rhizobium rosettiformans W3T
Rhizobium ipomoeae shin9-1T
Rhizobium sp. AG855
Rhizobium wuzhouense W44T
Rhizobium sp. MSSRF QS100
Ciceribacter lividus DSM 25528T
Ciceribacter sp. F8825
Rhizobium selenitireducens ATCC BAA-1503T
Rhizobium aggregatum complex
Agrobacterium
Neorhizobium sp. NCHU2750
Neorhizobium alkalisoli DSM 21826T
Neorhizobium huautlense DSM 21817T
Neorhizobium galegae HAMBI 540T
Pseudorhizobium
"R. oryzae/R. rhizosphaerae" group
Gellertiella hungarica DSM 29853T
Shinella
Pararhizobium antarcticum NAQVI 59T
Pararhizobium polonicum F5.1T
Pararhizobium giardinii H152T
Ensifer
Rhizobium
Mesorhizobium
0.02


a
total clusters = 10,501
cloud
shell
soft core
core
number of gene clusters
0
1000
2000
3000
4000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
number of genomes in clusters (occupancy)
b
soft core
(genomes ≥ 13),
3,775
core
(genomes = 14),
3,548
cloud
(genomes ≤ 2),
4,516
shell,
2,210
total gene clusters = 10,501
taxa = 14