Genomic analysis provides novel insights into diversification and taxonomy of Allorhizobium vitis (i.e. Agrobacterium vitis)

Nemanja Kuzmanović1,2*, Enrico Biondi3, Jörg Overmann4, Joanna Puławska5, Susanne Verbarg4, Kornelia Smalla1 and Florent Lassalle6,7,8*

Abstract
Background: Allorhizobium vitis (formerly named Agrobacterium vitis or Agrobacterium biovar 3) is the primary causal agent of crown gall disease of grapevine worldwide. We obtained and analyzed whole-genome sequences of diverse All. vitis strains to get insights into their diversification and taxonomy.

Results: Pairwise genome 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 several genomic species. Thus, we emended the description of All. vitis, which now refers to a restricted group of strains within the All. vitis species 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 current type strain of All. vitis, K309T. The type strain of All. ampelinum sp. nov. is S4T. We also identified sets of gene clusters specific to the All. vitis species complex, All. vitis sensu stricto and All. ampelinum, respectively, for which we predicted the biological function and infer the role in ecological diversification of these clades, including some we could experimentally validate. All. vitis species complex-specific genes confer tolerance to different stresses, including exposure to aromatic compounds. Similarly, All. vitis sensu stricto-specific genes confer the ability to degrade 4-hydroxyphenylacetate and a putative compound related to gentisic acid. All. ampelinum-specific genes have putative functions related to polyamine metabolism and nickel assimilation. Congruently with the genome-based classification, All. vitis sensu stricto and All. ampelinum were clearly delineated by MALDI-TOF MS analysis. Moreover, our genome-based analysis indicated that Allorhizobium is clearly separated from other genera of the family Rhizobiaceae.

Conclusions: Comparative genomics and phylogenomic analysis provided novel insights into the diversification and taxonomy of Allorhizobium vitis species complex, supporting our redefinition of All. vitis sensu stricto and description of All. ampelinum. Our pan-genome analyses suggest that these species have differentiated ecologies, each relying on specialized nutrient consumption or toxic compound degradation to adapt to their respective niche.
Keywords: Allorhizobium vitis sensu stricto, Allorhizobium ampelinum, Rhizobiaceae, Agrobacteria, Grapevine crown gall, Taxonomy, Plant pathogenic bacteria, Clade-specific genes, Ecological specialization, Pan-genome analysis

Background

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) [1]. This economically important plant disease may cause serious losses in nurseries and vineyards. All. vitis is widely distributed, 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 [1]). In one exceptional case, All. vitis was isolated from galls on the roots of kiwi in Japan [2].

All. vitis is an aerobic, non-spore-forming, Gram-negative, rod-shaped bacterium with peritrichous flagella [3]. 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 species complex) nor as biovar 2 (i.e., Rhizobium rhizogenes) [4]. Afterwards, several studies classified these atypical strains as Agrobacterium biovar 3 (biotype 3), based on their biochemical and physiological characteristics [5–7]. Serological analysis using monoclonal antibodies also allowed differentiation of Agrobacterium biovar 3 strains [8]. Polyphasic characterization involving DNA-DNA hybridization (DDH), phenotypic and serological tests clearly showed that Agrobacterium biovar 3 strains represent a separate species, for which the name Agrobacterium vitis was proposed [9]. However, multilocus sequence analysis (MLSA) suggested that A. vitis is phylogenetically distinct from the genus Agrobacterium, and prompted the transfer of this species to the revived genus Allorhizobium [10, 11]. The genus Allorhizobium was created by de Lajudie et al. [12] and initially included single species Allorhizobium undicola. Afterwards, Young et al. [13] proposed reclassification of All. undicola and its inclusion into the genus Rhizobium, while Costechareyre et al. [14] suggested that this species might belong to the genus Agrobacterium. However, these studies employed single gene phylogenies, which were insufficient to support such taxonomic revisions. The authenticity of the genus Allorhizobium and the clustering of All. vitis within it was unequivocally confirmed by genome-wide phylogenies [15, 16]. Moreover, distinctiveness of All. vitis with respect to the genus Agrobacterium was further supported by their different genome organization, with the genus Agrobacterium being characterized by the presence of a circular chromosome and a secondary linear chromid [17, 18]. Chromids are defined as large non-dispersable plasmids carrying essential functions [19]. In contrast to Agrobacterium, the All. vitis strains carry two circular chromosomes [18, 20, 21]. However, the smaller circular chromosome (named chromosome II) was later classified as a chromid in the fully sequenced strain All. vitis S4T [19]. 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 relies on the calculation of various overall genome relatedness indices (OGRIs) and estimation of genome-based phylogenies [22–24], largely replacing the traditionally used methods of 16S rRNA gene phylogeny and DDH [25, 26]. Genomic information were also highly recommended as essential for the description of new rhizobial and agrobacterial taxa [27]. 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 associated with the 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 [1]). In our previous study, we analyzed a representative collection of All. vitis strains originating from several European countries, Africa, North America, and Australia using MLSA, which indicated a high genetic diversity between strains, clustered into four main phylogenetic groups [28]. These data suggested that All. vitis might not be a homogenous species, but a species complex comprising several genomic species,
warranting further investigation of the diversification and evolution of *All. vitis* towards a more complete elucidation of its taxonomy.

In this work, we selected representative strains belonging predominantly to the two most frequent phylogenetic groups identified in our previous study [28] that included the well-studied *All. vitis* type strain K309\(^T\) and the fully sequenced strain S4\(^T\), respectively. We obtained draft genome sequences for 11 additional strains and performed comparative genomic and phylogenetic analyses to reveal the diversification history and synapomorphies 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 emending the description of *All. vitis* (*All. vitis sensu stricto*) and proposing the new species *All. ampelinum*.

**Results**

*Allorhizobium vitis* genome sequencing

Draft genome sequences were obtained for 11 *All. vitis* strains (Table 1), with average coverage depth ranging from 65- to 96- fold. The total size of draft genome assemblies ranged from 5.67 to 6.52 Mb, with a GC content ranging from 57.5–57.6% (Table 1), which was similar to the genomes of other *All. vitis* strains sequenced so far (Table S1b).

Core-genome phylogeny and overall genome relatedness indices measurements

A core-genome phylogeny was inferred for 14 strains of *All. vitis* (Table 1) and 55 reference *Rhizobiaceae* strains (Table S1a). A phylogenetic tree that was reconstructed from the concatenation of 344 non-recombining core marker genes confirmed the grouping of *Allorhizobium* species separately from other *Rhizobiaceae* genera (Figs. 1 and S1). The clade comprising all members of the genus *Allorhizobium* was well separated from its sister clade, which included members of the group provisionally named “*R. aggregatum* complex” [11], as well as representatives of the genus *Ciceribacter*.

*All. vitis* strains formed a well-delineated clade within the *Allorhizobium* genus (Figs. 1 and S1). Furthermore, *All. vitis* strains were clearly differentiated into two well-supported sub-clades (clades A and B), while strain Av2 branched separately from each of these two clades (Figs. 1 and S1). OGRIs values (Table S2) indicated that sub-clades A and B, as well as strain Av2, represent separate genomic species. In other words, the core-genome phylogeny and OGRI measurements showed that *All. vitis* is not a single species, but a species complex composed of at least three separate genomic species.

The first genomic species, corresponding to sub-clade A, comprises the type strain of *All. vitis* (strain K309\(^T\)) (Fig. 1). Although digital DDH (dDDH) values suggested that the cluster containing strains K309\(^T\) and KFB 253 might belong to a separate species compared to other strains comprised in this sub-clade (Table S2e), this was not supported by the other four OGRIs calculated here (Table S2a-d). Indeed, dDDH values for these strains (65.9–66.4%) were relatively close to the generally accepted threshold value of 70%. A revised description of the species *All. vitis*, hereafter referred to as *All. vitis sensu stricto*, is given below.

The second genomic species, corresponding to sub-clade B, included eight strains originating from various geographic areas (Table 1; Fig. 1). It included the well-studied strain S4\(^T\), whose high-quality genome sequence was described previously [18]. The dDDH 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 generally accepted for species delineation (Table S2e). However, other OGRIs unanimously indicated that strains from this sub-clade belong to the same species (Table S2a-d). A description of the novel species corresponding to sub-clade B, for which the name *Allorhizobium ampelinum* sp. nov. is proposed, is given below.

The third genomic species comprised strain Av2 alone (Figs. 1 and S1, Table S2). To get a more comprehensive insight into the diversity of the *All. vitis* species complex (AvSC), we conducted a second phylogenomic analysis where we included 34 additional genomes of *All. vitis* that were available in GenBank but not yet published (Table S1b). Based on core-genome phylogeny and average nucleotide identity (ANI) calculations (Fig. S2, Table S3), additional strains were taxonomically assigned as *All. vitis sensu stricto* (sub-clade A) and *All. ampelinum* (sub-clade B). Strain Av2 then grouped with three other strains originating from the USA (sub-clade D; Fig. S2). These four strains comprised in the sub-clade D were genetically very similar and exhibited >99.8 ANI between each other (Table S3). Moreover, additional sub-clades C and E were apparent, corresponding to two other new genomic species of the AvSC (Fig. S2, Table S3). Genomic species corresponding to sub-clade C and sub-clade D were closely related, as their ANI blast (ANIb) values were in the range from 94.62–94.93%, which is slightly below the threshold for species delimitation (~95–96%) [34].

Pan-genome analyses

A ML pan-genome phylogeny of the 64 *Rhizobiaceae* genome dataset was estimated from a matrix of the presence or absence of 33,396 orthologous gene clusters.
| Strain       | Species                | Geographic origin | Year of isolation | Reference | Genome sequencing | Contigs (N) | N50 (Kb) | Size (Mb) | GC Content (%) | Gene\(^a\) | CDS\(^a\) | Accession number       |
|-------------|------------------------|-------------------|-------------------|-----------|-------------------|-------------|---------|-----------|----------------|-----------|----------|------------------------|
| K309\(^T\) | All. vitis sensu stricto | Australia         | 1977              | [9]       | (51)              | 22          | 999     | 5.75      | 57.55          | 5188     | 5136     | LMVL0000000.2          |
| IPV-BO 6186 | All. vitis sensu stricto | Italy             | 2006              | [29]      | This study        | 79          | 608     | 5.80      | 57.57          | 5250     | 5196     | VOLK0000000.1          |
| IPV-BO 7105 | All. vitis sensu stricto | Italy             | 2007              | [28]      | This study        | 91          | 462     | 5.81      | 57.54          | 5214     | 5157     | VOLJ0000000.1          |
| KFB 239     | All. vitis sensu stricto | Serbia            | 2010              | [30]      | This study        | 82          | 456     | 6.15      | 57.57          | 5545     | 5490     | VOLF0000000.1          |
| KFB 253     | All. vitis sensu stricto | Serbia            | 2011              | [30]      | This study        | 70          | 401     | 5.81      | 57.56          | 5290     | 5235     | VOLF0000000.1          |
| S4\(^T\)    | All. ampelinum          | Hungary           | 1981              | [31]      | (18)              | CG\(^b\)   | CG\(^b\) | 6.32      | 57.47          | 5840     | 5770     | CP0000633.1-CP0000639.1 |
| IPV-BO 1861–5 | All. ampelinum         | Italy             | 1984              | [32]      | This study        | 110         | 971     | 5.67      | 57.63          | 5125     | 5072     | VOLM0000000.1          |
| IPV-BO 5159 | All. ampelinum          | Italy             | 2003              | [32]      | This study        | 160         | 269     | 6.47      | 57.56          | 5939     | 5883     | VOLL0000000.1          |
| KFB 243     | All. ampelinum          | Serbia            | 2011              | [30]      | This study        | 210         | 270     | 6.52      | 57.62          | 5963     | 5909     | VOLH0000000.1          |
| KFB 250     | All. ampelinum          | Serbia            | 2011              | [30]      | This study        | 135         | 356     | 6.47      | 57.60          | 5845     | 5790     | VOLG0000000.1          |
| KFB 254     | All. ampelinum          | Serbia            | 2011              | [30]      | This study        | 76          | 492     | 5.96      | 57.60          | 5433     | 5380     | VOLE0000000.1          |
| KFB 264     | All. ampelinum          | Serbia            | 2011              | [30]      | This study        | 75          | 714     | 5.92      | 57.52          | 5355     | 5299     | VOLL0000000.1          |
| V80/94      | All. ampelinum          | USA                | 1994              | [33]      | (92)              | 66          | 413     | 5.98      | 57.48          | 5467     | 5414     | NBZE0000000.1          |
| Av2         | Allorhizobium sp.       | Croatia           | 2006              | [28]      | This study        | 123         | 407     | 6.23      | 57.58          | 5713     | 5659     | VOLN0000000.1          |

\(^a\) Numbers based on Prokka annotations  
\(^b\) CG, Complete genome
The pan-genome phylogeny (Fig. 2; Fig. S3) presented the same resolved sub-clades of the *All. vitis* complex as the core-genome phylogeny (Fig. 1). Furthermore, *Rhizobiaceae* genera and clades were generally differentiated based on the pan-genome tree (Fig. 2; Fig. S3). Nevertheless, some inconsistencies were observed: tumorigenic strain *Neorhizobium* sp. NCHU2750 was more closely related to the representatives of the genus *Agrobacterium*, while nodulating *Pararhizobium gadii-nii* H152 was grouped with *Ensifer* spp. (Fig. 2; Fig. S3). These inconsistencies were also observed in another pan-genome phylogeny inferred using parsimony (data not shown). Such limitations of gene content-based phylogenies have previously been reported [35, 36].

Focusing on 14 AvSC strains, we identified 10,501 pan-genome gene clusters. The core-genome ('strict core' and 'soft core' compartments) of the species complex comprised 3,775 gene clusters (35.95% of total gene clusters), with 3,548 gene clusters strictly 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).

**Clade-specific gene clusters**

Homologous gene families specific to particular clades of interest, i.e. with contrasted presence pattern with respect to closely related clades, were identified using both Pantagruel or GET_HOMOLOGUES software packages. Both sets of inferred clade-specific genes were to a large extent congruent, although some differences were observed (Table S4), owing to the distinct approaches employed by these software packages [37, 38]. We focused on clusters of contiguous clade-specific genes for which we could predict putative molecular functions or association to a biological process. The results are summarized below and in Table S4.

**All. vitis species complex**

Based on Pantagruel and GET_HOMOLOGUES analyses, we identified 206 and 236 genes, respectively, that are specific to the *Av* species, i.e. present in all strains of *All. vitis sensu stricto, All. ampelinum* and *Allorhizobium* sp. Av2, and in no other *Allorhizobium* strain. AvSC-specific genes are mostly located on the second chromosome (chromid). While some AvSC-specific genes are found on the Ti plasmid and include the type 4 secretion system, this likely only reflects a sampling bias whereby all AvSC strains in our sample were tumorigenic and possessed a Ti plasmid. As such, Ti plasmid-encoded genes directly associated with pathogenicity were not further considered or discussed in this study.

Half of the AvSC-specific genes are gathered in contiguous clusters for most of which we could predict putative function (Table S4); most of the other half are...
Fig. 2 Maximum-likelihood pan-genome phylogeny of 69 strains belonging to the genus Allorhizobium and other Rhizobiaceae members (collapsed branches). The tree was estimated with IQ-TREE from the consensus (COGtriangles and OMCL clusters) gene presence/absence 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 IQ-TREE. 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 collapsed clades, is presented in the Figure S3.
scattered on chromosome 1 and have unknown function. Predicted functions of clustered genes revealed that they are strikingly convergent: most are involved in either environmental signal perception (four clusters), stress response (two clusters), aromatic compound and secondary metabolite biosynthesis (three clusters) and/or aromatic compound degradation response (two 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 or nucleotide uptake. Finally, one cluster on chromosome 1 encodes a putative autotransporter adhesin protein, which may have a role in plant commensalism and pathogenesis.

All studied AvSC strains carried a pehA gene encoding a polygalacturonase enzyme. Unlike other agrobacteria, All. vitis strains are known to produce a polygalacturonase, regardless of their tumorigenicity [39]. However, this gene was present also in All. taibaishanense 14971$^T$, All. terrae CC-HIH110$^T$ and All. oryziradicis N19$^T$, but absent in All. undicola ORS 992$^T$ and in other studied members of the Rhizobiaceae family.

Furthermore, we detected the presence of the gene encoding enzyme 1-aminocyclopropane-1-carboxylate deaminase (acdS) in all studied AvSC strains. This gene is considered to be important for plant-bacteria interaction through its involvement in lowering the level of ethylene produced by the plant [40]. We found this gene in all other Allorhizobium spp., and in some other Rhizobiaceae (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 findings [41].

Tartrate utilization ability was previously reported for most of the All. vitis strains [31, 42, 43]. Therefore, we searched AvSC genomes for the presence of tartrate utilization (TAR) regions. All strains except IPV-BO 6186 and IPV-BO 7105 carried TAR gene clusters. Moreover, we could not find any All. vitis-like TAR regions in any other Rhizobiaceae strain. Sequence comparison of TAR regions from AvSC strains using ANIb algorithm (Table S5) showed they could be divided into four types (Fig. S4). The first type is represented by a previously characterized TAR region called TAR-I, carried on the TAR plasmid pTrAB3 of strain AB3 [43, 44]. The second type included representatives of TAR-II (carried on pTiAB3) and TAR-III (carried on pTrAB4) regions, which were previously described to be related to each other [43, 45]. A third TAR region type, which we designate TAR-IV, was characterized by the absence of a second copy of ttuC gene (tartrate dehydrogenase). The TAR-IV region type is found in All. ampelinum strain S4$^T$, in which the TAR system is located on the large plasmid pAtS4c (initially named pTrS4) [44]. The TAR system of Allorhizobium sp. strain Av2 is a unique type (TAR-V), which is
related to region type TAR-I, but is characterized by the absence of the ttuA gene (a LysR-like regulator). We compared the distribution of these TAR region types in strain genomes, showing there is no TAR region type associated to any genomic species (Table S6). All. vitis sensu stricto strains K309T and KFB 253 carry a TAR-II/III region. In addition to TAR-II/III region, strain KFB 239 carries a TAR-I region (Table S6), a combination similar to that found in the well-characterized strain AB3 [43].

All. ampelinum strains S4T, IPV-BO 1856, and V80/94 contain a TAR-IV region, while the remaining All. ampelinum strains IPV-BO 5159, KFB 243, KFB 250 and KFB 254 additionally carry a TAR-II/III region (Table S6).

All. vitis sensu stricto

Using Pantagruel and GET_HOMOLOGUES pipelines, we identified 63 and 78 genes, that are specific to All. vitis sensu stricto (Av-specific, present in all five strains and in none of All. ampelinum), respectively. 32 of these Av-specific genes are clustered into four main loci in the genome of strain K309T, for which we could predict putative function (Table S4). One Av-specific gene cluster (Av-GC1, Table S4) comprised 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). Av-GC1 was located on Contig 1 (LMVL0200001.1) of reference strain K309T genome, which is likely part of the chromid, based on its high ANI with the chromid (Chromosome 2) of strain S4T, whose genome sequence is complete. BLAST searches showed that this gene cluster is also present in some representatives of Agrobacterium deltaense, i.e. Agrobacterium genospecies G7 (data not shown). Av-GC1 is predicted to 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 K309T and KFB 253 carrying a partial gene cluster, were phylogenetically separated and formed distinct sub-clades within All. vitis sensu stricto (Fig. 1).

Another Av-specific gene cluster (Av-GC4, Table S4) was annotated to be involved in the degradation of 4-hydroxyphenylacetate (MetaCyc pathway 3-HYDROXYPHENYLACETATE-DEGRADATION-PWY). Gene content and comparative analysis of the contig carrying this gene cluster suggested that Av-GC4 is carried on a putative plasmid of All. vitis sensu stricto (data not shown).

In addition, Av-specific gene clusters Av-GC2 and Av-GC3 (Table S4) were both predicted to be involved in amino-acid uptake and catabolism. However, we were not able to predict the precise molecular function of proteins and substrates of enzymes encoded by these Av-specific gene clusters. Both these gene clusters are likely located on a putative plasmid, as suggested by the presence of plasmid-related genes (replication- and/or conjugation-associated genes) on the same contigs.

All. ampelinum

Based on Pantagruel and GET_HOMOLOGUES analyses, we identified 97 and 128 genes, respectively, that are specific to All. ampelinum (Aa-specific, present in all eight strains and in none of All. vitis sensu stricto). Taking advantage of the finished status of strain S4T genome, we found that 52/97 specific genes identified by Pantagruel 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}). For 11 contiguous gene clusters we could predict putative function (Table S4). The Aa-specific gene clusters encode a variety of putative biological functions; an enrichment analysis of their functional annotations revealed a set of high-level biological processes that were over-represented: 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-GC10, which is located on the 631-kb megaplasmid pAtS4e and encodes the NikABCDE Ni^{2+} import system and a nickel-responsive transcriptional regulator NikR. Aa-GC10 additionally includes genes with predicted functions such as cation-binding proteins and a chaperone/thioredoxin, which may be involved in the biosynthesis of ion-associated cofactors.

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 bioinformatic analyses suggested that All. vitis sensu stricto strains carry a gene cluster encoding the
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 pH (~7.2 to ~6.5), indicating the production of acid from the substrate oxidation. On the other hand, All. ampelinum strains showed poor growth under culturing conditions, without change of pH.

Although All. vitis sensu stricto strains carry genes predicted to be involved in a degradation process of gentisic acid, this biochemical 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 AvSC 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 [30]. Taken together, all tested AvSC strains (Table 1) were able to produce alkali from tartrate. Interestingly, strains IPV-BO 6186 and IPV-BO 7105, for which we could not identify TAR gene clusters, were also positive for this test.

As a broader way to characterize and phenotypically distinguish strains, we used matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass-spectrometry (MS) of pure bacterial cultures. MALDI-TOF MS revealed diversity among the tested strains, while allowing to discriminate genomic species (Fig. S5).

**Relationship of the genus Allorhizobium and related Rhizobiaceae genera**

As indicated by the core-genome phylogeny, the genus *Allorhizobium* is clearly separated from the other representatives of the family *Rhizobiaceae*, including the "*R. aggregatum* complex", which, with the genus *Ciceribacter*, formed a well-delineated sister clade to *Allorhizobium* clade (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% average amino acid identity (AAI) among each other, and 70.79–72.63% AAI with members of the "*R. aggregatum* complex*/Ciceribacter* clade (Table S7). On the other hand, representatives of the genera *Shinella*, *Ensifer* and *Pararhizobium* showed 71.46–75.85% AAI between genera. Similarly, representatives of genera *Neorhizobium* and *Pseudorhizobium* showed 72.24–76.18% AAI between genera. In other words, AAI values suggested that the existing genera *Ensifer*, *Pararhizobium* and *Shinella*, or *Neorhizobium* and *Pseudorhizobium* were more closely related than the genus *Allorhizobium* and the "*R. aggregatum* complex*/Ciceribacter* clade. Genome-wide ANI (gANI) and percentage of conserved proteins (POCP) values similarly supported the divergence of the members of *Allorhizobium* genus and the "*R. aggregatum* complex*/Ciceribacter* clade (Table S7). Members of the genus *Allorhizobium* exhibited gANI and POCP values ranging from 73.55–76.86 and 55.27–66.17, respectively, when compared with members of the "*R. aggregatum* complex*/Ciceribacter* clade, values that were similar to these seen between representatives of the genera *Agrobacterium* and *Neorhizobium* (gANI 74.66–77.45; POCP 59.96–65.58).

**Discussion**

*Allorhizobium vitis* is not a single species

Genomic analyses allowed us to unravel the substantial taxonomic diversity within *All. vitis*. In particular, whole-genome sequence comparisons and phylogenomic analyses clearly showed that *All. vitis* is not a single species, but represents a species complex composed of several genomic species. Similarly, *Agrobacterium* biovar 1 (i.e. *A. tumefaciens*) was initially considered a single species, but was later designated as a species complex comprising closely related, but distinct genomic species. Several studies applying DDH initially demonstrated this species diversity within *Agrobacterium* biovar 1 [46–48], which was later supported by results obtained with AFLP [49, 50], housekeeping gene analysis [10, 11, 14] and whole-genome sequence analysis [35]. Although Ophel and Kerr [9] also performed DDH for several *All. vitis* strains, diversity within this species remained unknown because these authors only studied strains that belonged to *All. vitis sensu stricto* as defined here.

Our previous study based on the analysis of several housekeeping gene sequences suggested the existence of several phylogenetic groups within *AvSC* [28]. The present study focused on two phylogenetic groups defined in our previous study: the first comprises the type strain of *All. vitis* (strain K309T) [9, 51], whereas the second includes the well-characterized and completely sequenced strain S4T [18]. Consequently, we amended the description of *All. vitis*, which now refers to the limited group within *AvSC* strains (*All. vitis sensu stricto*) and proposed a description of a novel species, *All. ampelinum* sp. nov. (see formal description below).

As indicated by the genome analysis of a larger set of strains available from the NCBI GenBank database, the taxonomic diversity of *AvSC* is not limited to *All. vitis sensu stricto* and *All. ampelinum* sp. nov. However, the description of sub-clades C, D and E (Fig. S2) as separate species was considered outside the scope of this study, because the sequencing of these strains was not conducted by our group and their draft genome sequences are yet to be described in scientific publication(s). In addition, it is not clear whether sub-clades C and D represent a single or separate species. Further comprehensive genomic analysis of diverse members of these clades is required to elucidate relationships between them.

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

The convergence of functions encoded by the *AvSC*-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 AvSC-specific gene set indicates that adaptation to a changing environment is to be a key feature of their ecology.

We also searched genomes of AvSC 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 [39, 52, 53], and tartrate degradation [42] were proposed to contribute to the specialization of All. vitis to its grapevine host. In addition, polygalacturonase activity might be involved in the process of the invasion of the host plant, as postulated previously for other rhizobia [54]. Although all AvSC strains carried the pehA gene encoding a polygalacturonase enzyme, this gene was not restricted to this bacterial group, as it was also present in all other Allorhizobium spp. strains included in our analysis, except for All. undicola.

All AvSC 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, we speculate that strains IPV-BO 6186 and IPV-BO 7105 must carry another type of TAR system, distinct from those described so far in other All. vitis strains. Furthermore, some diversity between TAR regions and variable distribution patterns of different TAR regions among strains were observed, in line with previously reported data [43]. The existence of non-tartrate-utilizing strains was also documented in the literature [43]. Considering the fact that tartrate utilization in All. vitis has only been observed as plasmid-borne [44, 45, 55], this suggests that tartrate utilization is an accessory trait that can be readily gained via the acquisition of a plasmid encoding this trait and selected for in tartrate-abundant environments. Because grapevine is rich in tartrate [56], utilization of this substrate may enhance the competitiveness of AvSC strains in colonizing this plant species [42].

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 replications may be an important part of these species’ adaptive core-genome, as previously observed in the A. tumefaciens species complex [35]. Ecological differentiation of the two main species of the AvSC 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 of which the predicted function was the degradation of gentisic acid, we could not experimentally demonstrate this trait. Gentiisic acid was detected in grapevine leaves [57] and is likely present in other parts of this plant. This compound was reported as a plant defense signal that can accumulate in some plants responding to compatible viral pathogens [58, 59]. In addition, a sub-clade within All. vitis sensu stricto composed of strains K3097 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 [60]. 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 predicted, and demonstrated that all studied All. vitis sensu stricto strains have the specific ability to degrade 4-hydroxyphenylacetate, an activity that may contribute to the detoxication of aromatic compounds and thus to the survival of this bacterium in soil, notably in competition against bacteria lacking this pathway.

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 Escherichia coli [61]. Hydrogenase function has in turn been proposed as a potential mechanism for detoxication of phenolic compounds in A. vitis [62] and may thus have an important role in survival in the rhizosphere.

Delineation of the genus Allorhizobium

The genus Allorhizobium was clearly differentiated from other Rhizobiaceae genera based on core- and pan-genome-based phylogenies, in line with previous studies employing genome-wide phylogeny [15, 16]. We included diverse AvSC strains into our analysis, confirming 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”/Ciceribacter clade is still unresolved. Although MLSA suggested that “R. aggregatum complex” is a sister clade of the genus Agrobacterium [11], the more thorough phylogenetic analyses performed in this study rather showed that the “R. aggregatum complex” grouped with Ciceribacter spp., in a clade that is more closely related to the genus Allorhizobium. Presently, there are no widely accepted criteria and scientific consensus regarding the delineation of new bacterial genera [27]. In this study, existing Rhizobiaceae genera were compared using several delineation methods proposed in the literature, such as AAI [63, 64], POCP [65], or gANI and alignment fraction (AF) [66], which we
complemented with genome-based phylogenies. Taken together, our genome-based analysis suggested that *Allorhizobium* represents a genus clearly separated from other *Rhzobiaceae* genera, including closely related “*R. aggregatum* complex”/Ciceribacter clade. A separate and more focused analysis is, however, required to explore the taxonomic diversity and structure of the “*R. aggregatum* complex”/Ciceribacter clade.

**Conclusions**

Whole-genome sequence comparisons and phylogenomic analyses classified *All. vittis* strains within the genus *Allorhizobium*, which was clearly differentiated from other *Rhzobiaceae* genera, including the closely related “*R. aggregatum* complex”/Ciceribacter clade. We revealed an extensive and structured genomic diversity within *All. vittis*, which in fact represents a species complex composed of several genomic species. Consequently, we emended the description of *All. vittis*, now encompassing a restricted group of strains within the *AvSC* (i.e. *All. vittis sensu stricto*) and proposed a description of a novel species, *All. ampelinum* sp. nov. Further analyses including pan-genome reconstruction and phylogeny-driven comparative genomics revealed loci of genomic differentiation between these two species. Functional analysis of these species-specific loci suggested that these species are ecologically differentiated as they can consume specific nutrient sources (*All. ampelinum*), or degrade specific toxic compounds (*All. vittis sensu stricto*). We identified another two potential genomic species within the *AvSC*, further characterization of which was prevented by the limited diversity of available isolates. We also described how accessory genomic regions associated with the colonization of grapevine host plant are distributed across species, and how they combine to form diverse genotypes. However, given the complete bias in sampling of *All. vittis* strains – all grapevine pathogens – the ecological significance of this genetic diversity remains unclear. We encourage future studies to integrate genomic data from new genomically diverse isolates, to further unravel the ecological basis of *AvSC* diversification.

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

The description of *Agrobacterium vittis* is provided by Ophel and Kerr [9]. Young et al. [13] proposed the transfer of *A. vittis* to the genus *Rhizobium*, but it was neither widely accepted by the scientific community nor supported by further studies [14, 67, 68]. Mousavi et al. [11] reclassified this species to the genus *Allorhizobium*, which was included into the Validation list no. 172 of the IJSEM [69]. Hördt et al. [16] emended a description of *All. vittis* by including genome sequence data for its type strain, which was published in the List of changes in taxonomic opinion no. 32 [70].

As shown in this study, *All. vittis sensu stricto* includes a limited group of strains that can be differentiated from other *All. vittis* genomic species and other *Allorhizobium* species based on OGRIs, such as ANI, as well as by core-genome phylogeny. Moreover, *All. vittis sensu stricto* can be differentiated from other species of *AvSC* by analysis of sequences of housekeeping genes *dnaK*, *gyrB* and *recA* [28]. Finally, this study demonstrated that strains belonging to this species can be distinguished from *All. ampelinum* by MALDI-TOF MS analysis. Unlike any *All. ampelinum*, all tested *All. vittis sensu stricto* strains are able to produce acid in a medium containing 4-hydroxyphenylacetate. However, this apparently species-specific trait is borne by a plasmid, and could possibly be transmitted to closely related species.

The whole-genome sequence of type strain K309T is available in GenBank under the accessions LMYL00000000.2 and GCA_001541345.2 for the Nucleotide and Assembly databases, respectively [51]. The genomic G+C content of the type strain is 57.55%. Its approximate genome size is 5.75 Mbp.

Basonym: *Agrobacterium vittis* Ophel and Kerr 1990.

The type strain, K309T (= NCPPB 3554T = HAMBI 1817T = ATCC 49767T = CIP 105853T = ICMP 10752T = IFO 15140T = JCM 21033T = LMG 8750T = NBRC 15140T), 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 the protologue (Table 2).

*All. ampelinum* (am.pe.li’num. Gr. n. ampelos grape-vine; Gr. adj. ampelinos and N.L. neut. adj. ampelinum of the vine).

*All. ampelinum* strains were formerly classified in the species *All. vittis*. However, our genomic data showed that they can be distinguished from *All. vittis sensu stricto* and other *All. vittis* genomic species based on OGRIs (e.g. ANI and dDHH) and core-genome phylogeny, as well as by analysis of sequences of housekeeping genes [28]. Furthermore, *All. ampelinum* can be differentiated from *All. vittis sensu stricto* by MALDI-TOF MS analysis.

The type strain, S4T (= DSM 112012T = ATCC BAA-846T) was isolated from grapevine tumor in Hungary in 1981.

**Methods**

*Allorhizobium vittis* strains

*All. vittis* 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 [28].

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 2 × 487,883 – 2 × 2,309,377 paired reads per genome. Trimming and quality filtering of raw reads were conducted using Trimmomatic (Galaxy Version 0.36.5) [71] implemented on the Galaxy Web server [72]. 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 2 × 1,037,619 – 2 × 1,443,575 paired reads. Demultiplexing and adapter clipping was done using the bcl2fastq2 conversion software (Illumina, USA).

Genome assembly and annotation
De novo genome assemblies were performed using the SPAdes genome assembler (Galaxy Version 3.12.0 + galaxy1) [73]. For genomes sequenced on the MiSeq and NextSeq platforms, both sets of reads were conducted for assembly. The genome sequences were deposited to DDBJ/ENA/GenBank under the Whole Genome Shotgun projects accession numbers listed in Table 1, under BioProject ID PRJNA557463.

The genome sequences were annotated using Prokka (Galaxy Version 1.13) [74] and NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) [75]. Prokka Version 1.14.6 was used to annotate genomes as a part of the Pantagruel pipeline (task 0; see below and Supplementary Methods). Functional annotation of proteins encoded by each gene family clustered by Pantagruel was conducted by the InterProScan software package Version 5.42–78.0 [76] as implemented in the Pantagruel pipeline (Task 4). Additionally, annotation of particular sequences of interest and metabolic pathway prediction were performed using BlastKOALA and GhostKOALA (last accessed in December, 2020) [77]. Protein sequences analyzed were subjected to Pfam domain searches (database release 32.0, September 2018, 17,929 entries) [78]. Metabolic pathway prediction was performed using KEGG [79] and MetaCyc [80] databases (last accessed in December, 2020).

The NCBI BLASTN and BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi), as well as BLAST search tool of KEGG database (last accessed in December, 2020) [79], were used for ad-hoc sequence comparisons at the nucleotide and amino acid levels, respectively.

Core- and pan-genome phylogenomic analyses
For phylogenomic analyses, whole genome sequences of 69 Rhizobiaceae strains were used, including 14 strains of All. vitis (Table 1) and 55 reference Rhizobiaceae strains (Table S1a). Additionally, in order to further explore the phylogenetic diversity of All. vitis, another core-genome phylogeny was inferred from an extended dataset that also included 34 All. vitis genomes available from GenBank but not yet published in peer-review journals by sequence depositors (Table S1b). To build phylogenies based on the core-genome (supermatrix of concatenated non-recombining core gene alignments) and on the pan-genome (homologous gene cluster presence/absence matrix), we used the GET_HOMOLOGUES Version 10,032,020 [38] and GET_PHYLOMARKERS Version 2.2.8.1_16Jul2019 [81] software packages. Details of the bioinformatic pipeline and used options are described in the Supplementary Methods.

Overall genome relatedness indices
To differentiate between the strains, different OGRIs were computed. For species delimitation, we relied on the values of ANI [34, 82] and dDDH [83] among strain genomes. Because different implementations of the ANI metric are known to give slightly different results [84], ANI was calculated using several programs: PyANI Version 0.2.9 (for metrics ANIb and ANIm) [85] (https://github.com/widdowquinn/pyani), OrthoANIu Version 1.2 [86] and FastANI Version 1.2 [87] tools. dDDH values were calculated using the Genome-to-Genome Distance Calculator (GGDC) Version 2.1 [83].

For genus delimitation, we relied on AAI [22, 63, 82], gANI and AF [88], and POCP [65]. AAI values were calculated with CompareM Version 0.0.23 (https://github.com/dparks1134/CompareM). gANI and AF values were obtained by the ANIcalculator Version 1.0 [88]. POCP values were calculated using GET_HOMOLOGUES software package [38]. Details of the used software and options are given in the Supplementary Methods.
Genome gene content analyses and identification of clade-specific genes

To explore the distribution of genome gene contents, we conducted further pan-genome analyses on more focused datasets, using two different bioinformatics pipelines, from which we present a consensus. Firstly, a pan-genome database was constructed using the Pangenegruel pipeline Version 00aaac71f85a2afa164949b86fb-c5b1613556f36 under the default settings as described previously [36, 37] and in Supplementary Methods. Because of computationally intensive tasks undertaken in this pipeline, the dataset was limited to the Allorhizobium genus and its sister clade “Rhizobium aggregatum complex”/Ciceribacter (28 strains).

Secondly, we analyzed a more focused dataset comprised of the 14 AvSC strains (Table 1) and four Allorhizobium spp. (All. oryzaeviscidis N19T, All. taibaishanense 14971T, All. terrae CC-HIH110T and All. undicola ORS 992T; Table S1a), using the GET_HOMOLOGUES software package [38]. Pan-genome gene clusters were classified into core, soft core, cloud and shell compartments [89] and species-specific gene families were identified from the pan-genome matrix. For details on the used scripts and options, see Supplementary Methods.

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, France). Utilization of sole carbon sources was tested with Biolog GEN III microplates using protocol A, according to the instructions of the manufacturer (Biolog, Inc., USA).

The metabolism of 4-hydroxyphenylacetic acid (p-hydroxyphenylacetic acid; Acros Organics, Product code: 121,710,250) and gentisic acid (2,5-dihydroxybenzoic acid; Merck, Product Number: 841745) was performed in AT minimal medium [90, 91] 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 acids 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 [5].

MALDI-TOF Mass Spectrometry analysis

Sample preparation for MALDI-TOF MS was carried out according to the Protocol 3 described by Schumann and Maier [92]. Instrument settings for the measurements were as described previously by Tóth et al. [93].

The dendrogram was created using the MALDI Biotyper Compass Explorer software (Bruker, Version 4.1.90).

Abbreviations

AAI: Average amino acid identity; AF: Alignment fraction; ANI: Average nucleotide identity; DDH: DNA-DNA hybridization; dDDH: Digital DDH; gANI: Genome-wide ANI; MS: Mass-spectrometry; OGI: Overall genome relatedness index; MLSA: Multilocus sequence analysis; PGAP: Prokaryotic genomes annotation pipeline; POC: Percentage of conserved proteins; TAR Tartrate utilization; Ti Tumor-inducing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08662-x.

Additional file 1: Fig. S1. Maximum-likelihood core-genome phylogeny of 69 strains belonging to the genus Allorhizobium and other Rhizobiaceae members (uncollapsed). 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 IQ-TREE. 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 collapsed clades, is presented in Figure 1.

Additional file 2: Fig. S2. Maximum-likelihood core-genome phylogeny of 103 strains belonging to the genus Allorhizobium (including 34 additional strains of All. vitis species complex strains whose sequences are available in GenBank but not associated to a published study) and other Rhizobiaceae members. 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 IQ-TREE. 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 in the top-right corner represents the distribution of ANIb values for genomic sequences of the clade corresponding to the All. vitis species complex, relative to the typical species delimitation threshold of 95%.

Additional file 3: Fig. S3. Maximum-likelihood pan-genome phylogeny of 69 strains belonging to the genus Allorhizobium and other Rhizobiaceae members (uncollapsed). The tree was estimated with IQ-TREE from the concatenated alignment of 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 IQ-TREE. 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+F+ASC+R5 model. The same tree, but with collapsed clades, is presented in Figure 2.

Additional file 4: Fig. S4. Heatmap representation of the average nucleotide identity (ANIb) for TAR regions of All. vitis species complex strains. PyANI program Version 0.2.9 (https://github.com/widdowquinn/pyani) was used to calculate ANIb values and generate the clustered heatmap.

Additional file 5: Fig. S5. Score-oriented dendrogram showing the similarity of the MALDI-TOF mass spectra of 14 All. vitis species complex strains studied. The dendrogram was created using the MALDI Biotyper Compass Explorer software (Bruker, Version 4.1.90).

Additional file 6: Table S1. List of additional strains and GenBank/EMBL/DDBJ accession numbers for their nucleotide sequences used in this study. a) List of 55 reference Rhizobiaceae strains and GenBank/EMBL/DDBJ accession numbers for their nucleotide sequences used in
this study. b) List of additional 34 *A. vitis* species complex strains and GenBank/ENA/BMBl accession numbers for their nucleotide sequences used in this study. Although available in the public nucleotide sequence databases, these genome sequences have not yet been presented in peer-reviewed study by sequence depositors.

Additional file 7: Table S2. Pairwise OGRI comparisons amongst 14 *A. vitis* species complex strain genomes towards species delimitation. a) ANlb comparisons. b) ANNm comparisons. c) orthoANnu comparisons. d) fastANl comparisons. e) DDBJ comparisons.

Additional file 8: Table S3. Pairwise ANlb comparisons amongst extended set of *A. vitis* species complex strain genomes towards species delimitation. Additionally, reference *Rhizobiaceae* strains were also included.

Additional file 9: Table S4. Clusters of contiguous clade-specific genes. Clusters were identified amongst sets of genes deemed specific of the focal clade based on detection by either Pantagruel or GET_HOMOLOGUES pipelines. a) Clusters of genes specific to *A. vitis* species complex (present in all *A. vitis* sensu stricto, *A. amplexicaulis* and *A. hoffmannii* sp. Av2, and in no other *A. hoffmannii* spp.). b) Clusters of genes specific to *A. vitis sensu stricto* (present in all five tested strains and in none of *A. amplexicaulis*). c) Clusters of genes specific to *A. amplexicaulis* (present in all eight tested strains and in none of *A. vitis sensu stricto*).

Additional file 10: Table S5. Pairwise ANlb values between tartrate utilization (TAR) regions of *A. vitis* species complex strains.

Additional file 11: Table S6. Tartrate utilization (TAR) region genotype and tartrate metabolism phenotype (production of alkali from L-tartaric acid) of *A. vitis* species complex strains.

Additional file 12: Table S7. Pairwise OGRI comparisons amongst 69 *Rhizobiaceae* strain genomes towards genus delimitation. a) Average amino acid identity (AAI). b) percentage of conserved proteins (POCP). c) genome-wide average nucleotide identity (gANI) and alignment fraction (AF), with AF values indicated in parentheses.

Additional file 13. Supplementary methods.

Acknowledgements
We would like to thank Cathrin Spröer and Boyke Bunk for conducting the Illumina sequencing. We are grateful to Anja Frühling and Ulink Steiner for support in phenotypic tests.

Authors’ contributions
NK and FL conceived and designed the study, and analyzed data. FL conceived and implemented the bioinformatic analysis pipeline Pantagruel. SV received and implemented the bioinformatic analysis pipeline Pantagruel. SV, NK and FL wrote the manuscript. All authors read, discussed, edited and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This research was supported by the Georg Forster Fellowship for postdoctoral research from the Alexander von Humboldt-Foundation, Bonn, Germany. The work of FL was supported by a Medical Research Council (MRC) grant (MR/N010760/1) and the Wellcome Trust Grant [206194]. 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).

Availability of data and materials
The genome sequences generated in this study were deposited in DDBJ/ENA/GenBank under the Whole Genome Shotgun projects accession numbers listed in Table 1, under BioProject ID PRJNA557463. The versions described in this paper are first versions. All other relevant data (including output of analyses) referring to this project have been deposited on Figshare under the project accession 20,894, available at figshare [https://figshare.com/], with individual items accessible at DOIs: https://doi.org/10.6084/m9.figshare.17105267, https://doi.org/10.6084/

References
1. Kuzmanović N, PulvArska J, Hao L, Burr TJ. The ecology of Agrobacterium vitis and management of crown gall disease in vineyards. Curr Top Microbiol Immunol. 2018;418. Springer, Cham. https://doi.org/10.1007/82_2018_85.
2. Sawada H, leki H. Crown gall of kiwi caused by *Agrobacterium tumefaciens* in Japan. Plant Dis. 1992;76:212.
3. Young JM, Kerr A, Sawada H. *Genus Agrobacterium* Conn 1942, 359p. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. Bergey’s Manual of Systematics of Archaea and Bacteria. The Proteobacteria. 2ed ed. New York: Springer-Verlag. 2005. p. 340–5.
4. Panagopoulos CG, Psallidas PG. Characteristics of Greek isolates of *Agrobacterium tumefaciens*. (E.F. Smith & Townsend) Conn. J Appl Bacterial. 1973;36:233–40.
5. Kerr A, Panagopoulos CG. Biotypes of Agrobacterium radiobacter var. *tumefaciens* and their biological control. J Phytopathol. 1977;90(2):172–9.
6. Panagopoulos CG, Psallidas PG, Alivizatos AS, editors. Studies on biotype 3 of *Agrobacterium radiobacter* var. *tumefaciens*. 4th International conference on plant pathogenic bacteria; 1978; Angers, France.
7. Süle S. *Biotypes of Agrobacterium tumefaciens* in Hungary. J Appl Bacteriol. 1978;43(4):207–13.
8. Bishop AL, Burr TJ, Mittak VL, Katz BH. A monoclonal antibody specific to *Agrobacterium tumefaciens* biovar 3 and its utilization for indexing grapevine propagation material. Phytopathol. 1989;79:995–8.
9. Ophel K, Kerr A. *Agrobacterium vitis* sp. nov. for strains of *Agrobacterium biovar 3* from grapevines. Int J Syst Bacteriol. 1990;40(3):236–41.
10. Mousavi SA, Ostermann J, Hallberg N, Nesme X, Lavire C, Vial L, et al. Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. Syst Appl Microbiol. 2014;37(3):208–15.
11. Mousavi SA, Willems A, Nesme X, de Lajudie P, Lindström K. Revised phylogeny of *Rhizobiaceae*: Proposal of the delineation of *Pararhizobium*
gen. nov, and 13 new species combinations. Syst Appl Microbiol. 2015;38:84–90.

12. de Lajudie P, Laurent-Fulele E, Willems A, Torek U, Coopman R, Collins MD, et al. Allorhizobium undulcuta gen. nov, sp. nov, nitrogen-fixing bacteria that efficiently nodulate Neptunia natans in Senegal. Int J Syst Evol Microbiol. 1998;48(4):1277–90.

13. Young JM, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H, 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 undulcuta de Lajudie et al. 1998 as new combinations: Rhizobium radiobacter, Rhizobium, R. rubi, R. undulcuta and R. vitis. Int J Syst Evol Microbiol. 2001;51(Pt1):189–103.

14. Costechareyre D, Rhouma A, Lavire C, Portier P, Chapulliot D, Bertolla F, et al. Rapid and efficient identification of Agrobacterium species by recA allele analysis. Agrobacterium recA diversity. Microb Ecol. 2010;60(4):862–72.

15. Ormeño-Orrillo E, Servín-Garcidueñas LE, Rangel MA, González V, Peralta H, Mora J, et al. Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. Syst Appl Microbiol. 2015;38(4):287–91.

16. Hördt A, López MG, Meier-Kolthoff JP, Schleuning M, Weinhold L-M, Ramírez-Bahena MH, Vial L, Lassalle F, Diel B, Chapulliot D, Daubin V, et al. Isothermal amplification and lateral-flow assay for detecting crown-gall causing Agrobacterium spp. Phytopathol. 2017;107(9):1062–8.

17. Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA. 2009;106(45):19126–31.

18. Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, et al. Genome sequences of three Alphaproteobacteria + Agrobacterium supercluster characterized by the presence of a linear chromid. Mol Phylogenet Evol. 2020;11(468). https://doi.org/10.3389/fmicb.2020.00468.

19. RAMíREZ-BALIENTES MH, ViÁ, L, Lassalle F, Diel B, Chapulliot D, Daubin V, et al. 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. Mol Phylogenet Evol. 2014;73:202–7.

20. Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, et al. Genome sequences of three Agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. J Bacteriol. 2009;191(8):2501–11.

21. Harrison PW, Lower RP, Kim NK, Young JP. Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. Trends Microbiol. 2010;18(4):141–8.

22. Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allardet-Servent J, Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005;187(18):6258–64.

23. Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol. 2014;64(2):316–24.

24. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardised bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36(10):996–1004.

25. Wayne LG, Brenner DJ, Colwell RR, Grimont PDa, Kandler O, Krichevsky MI, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol. 1987;37:463–4.

26. Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol. 1994;44:846–9.

27. de Lajudie PM, Andrews M, Ardley J, Eardly B, Jumas-Bilak E, Kuzmanovic N, et al. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. Int J Syst Evol Microbiol. 2019;69(7):1852–63.

28. Kuzmanovic N, Biondi E, Bertaccini A, Obradović A. Genetic relatedness and recombination analysis of Allorhizobium vitis strains associated with grapevine crown gall outbreaks in Europe. J Appl Microbiol. 2013;115:199–205.

29. Kuzmanovic N, Ivanovic M, Prokic A, Galic K, Zlatkovic N, Obradovic A. Characterization and phylogenetic diversity of Agrobacterium vitis from Serbia based on sequence analysis of 16S–23S rRNA internal transcribed spacer (ITS) region. Eur J Plant Pathol. 2014;140:757–68.

30. Kuzmanovic N, Ivanovic M, Prokic A, Galic K, Zlatkovic N, Obradovic A. Characterization and phylogenetic diversity of Agrobacterium vitis from Serbia based on sequence analysis of 16S–23S rRNA internal transcribed spacer (ITS) region. Eur J Plant Pathol. 2014;140:757–68.
