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

Nemanja Kuzmanović[1,2*], Enrico Biondi[3], Jörg Overmann[4], Joanna Puławska[5], Susanne Verbarg[4], Kornelia Smalla[1], Florent Lassalle[6,7*]

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

[2]Julius Kühn Institute, Federal Research Centre for Cultivated Plants (JKI), Institute for Plant Protection in Horticulture and Forests, Messeweg 11-12, 38104 Braunschweig, Germany (current affiliation)

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

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

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

[6]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

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

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

Abstract

Background: *Allorhizobium vitis* (formerly named *Agrobacterium vitis* or *Agrobacterium* biovar 3) is the primary causative 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*, K309^T. The type strain of *All. ampelinum* sp. nov. is S4^T. 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

69 **Background**

70 *Allorhizobium vitis* (formerly named *Agrobacterium vitis* or *Agrobacterium* biovar 3) is a
71 bacterium primarily known as a plant pathogen causing crown gall disease of grapevine (*Vitis*
72 *vinifera*) (1). This economically important plant disease may cause serious losses in nurseries
73 and vineyards. *All. vitis* is widely distributed pathogen, detected in almost all grapevine
74 growing regions throughout the world. This bacterium seems to be associated almost
75 exclusively with grapevine. It has been isolated from crown gall tumors, xylem sap, roots,
76 rhizosphere, non-rhizosphere soil of infected vineyards, decaying grape roots and canes in
77 soil, but also from the phyllosphere of grapevine plants (reviewed in (1)). In one exceptional
78 case, *All. vitis* was isolated from galls on the roots of kiwi in Japan (2).

79 *All. vitis* is an aerobic, non-spore-forming, Gram-negative, rod-shaped bacterium with
80 peritrichous flagella (3). It is a member of the alphaproteobacterial family *Rhizobiaceae*,
81 together with other genera hosting tumor-inducing plant pathogens, including
82 *Agrobacterium* and *Rhizobium*. With time, the taxonomy of *All. vitis* has undergone various
83 changes. Tumorigenic strains associated with crown gall of grapevine were initially defined
84 as an atypical group that could neither be classified as *Agrobacterium* biovar 1 (i.e.,
85 *Agrobacterium tumefaciens* species complex) nor as biovar 2 (i.e. *Rhizobium rhizogenes*) (4).
86 Afterwards, several studies classified these atypical strains as *Agrobacterium* biovar 3
87 (biotype 3), based on their biochemical and physiological characteristics (5-7). Serological
88 analysis using monoclonal antibodies also allowed differentiation of *Agrobacterium* biovar 3
89 strains (8). Polyphasic characterization involving DNA-DNA hybridization (DDH), phenotypic
90 and serological tests clearly showed that *Agrobacterium* biovar 3 strains represent a
91 separate species, for which the name *Agrobacterium vitis* was proposed (9). However, multi-
92 locus sequence analysis (MLSA) suggested that *A. vitis* is phylogenetically distinct from the

93 genus *Agrobacterium*, and prompted the transfer of this species to the revived genus
94 *Allorhizobium* (10, 11).

95 The genus *Allorhizobium* was created by de Lajudie et al. (12) and initially included single
96 species *Allorhizobium undicola*. Afterwards, Young et al. (13) proposed reclassification of *All.*
97 *undicola* and its inclusion into the genus *Rhizobium*, while Costechareyre et al. (14)
98 suggested that this species might belong to the genus *Agrobacterium*. However, these
99 studies employed single gene phylogenies, which were insufficient to support such
100 taxonomic revisions. The authenticity of the genus *Allorhizobium* and the clustering of *All.*
101 *vitis* within it was unequivocally confirmed by genome-wide phylogenies (15, 16). Moreover,
102 distinctiveness of *All. vitis* with respect to the genus *Agrobacterium* was further supported
103 by their different genome organization, with the genus *Agrobacterium* being characterized
104 by the presence of a circular chromosome and a secondary linear chromid (17, 18). Chromids
105 are defined as large non-dispensable plasmids carrying essential functions (19). In contrast
106 to *Agrobacterium*, the *All. vitis* strains carry two circular chromosomes (18, 20, 21).
107 However, the smaller circular chromosome (named chromosome II) was later classified as a
108 chromid in the fully sequenced strain *All. vitis* S4^T (19). Additionally, genomes of *All. vitis* and
109 other agrobacteria include a variable number of plasmids.

110 In recent years, genomics has significantly impacted the taxonomy of bacteria, leading to
111 the revisions in classification of different bacterial taxa. In particular, a novel genomics-based
112 taxonomy primarily relies on the calculation of various overall genome relatedness indices
113 (OGRIs) and estimation of genome-based phylogenies (22-24), largely replacing the
114 traditionally used methods of 16S rRNA gene phylogeny and DDH (25, 26). Genomic
115 information were also highly recommended as essential for the description of new rhizobial
116 and agrobacterial taxa (27). In addition, it has been recommended that some functions and

117 phenotypic characters may not be considered for taxonomic classification. This particularly
118 applies to the tumor-inducing ability of agrobacteria, which is mainly associated with the
119 dispensable tumor-inducing (Ti) plasmid.

120 Information on genetic diversity and relatedness of strains responsible for crown gall
121 disease outbreaks provide important insights into the epidemiology, ecology and evolution
122 of the pathogen. Numerous studies indicated that *All. vitis* strains are genetically very
123 diverse (reviewed in (1)). In our previous study, we analyzed a representative collection of
124 *All. vitis* strains originating from several European countries, Africa, North America, and
125 Australia using MLSA, which indicated a high genetic diversity between strains, clustered
126 into four main phylogenetic groups (28). These data suggested that *All. vitis* might not be a
127 homogenous species, but a species complex comprising several genomic species, warranting
128 further investigation of the diversification and evolution of *All. vitis* towards a more
129 complete elucidation of its taxonomy.

130 In this work, we selected representative strains belonging predominantly to the two most
131 frequent phylogenetic groups identified in our previous study (28) that included the well-
132 studied *All. vitis* type strain K309^T and the fully sequenced strain S4^T, respectively. We
133 obtained draft genome sequences for 11 additional strains and performed comparative
134 genomic and phylogenetic analyses to reveal the diversification history and synapomorphies
135 of these groups. In parallel, we investigated phenotypic features of selected strains. The
136 combination of these approaches allows us to revise the taxonomy within this group,
137 notably by emending the description of *All. vitis* (*All. vitis sensu stricto*) and proposing the
138 new species *All. ampelinum*.

139

140 **Results**

141 ***Allorhizobium vitis* genome sequencing**

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

146

147 **Table 1**

148

149 **Core-genome phylogeny and overall genome relatedness indexes measurements**

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

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

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

171 The second genomic species, corresponding to sub-clade B, included eight strains
172 originating from various geographic areas (Table 1; Fig. 1). It included the well-studied strain
173 S4^T, whose high-quality genome sequence was described previously (18). The dDDH value
174 obtained from the comparison of strain KFB 254 with strain IPV-BO 1861-5 was below, but
175 very close to the 70 % threshold value generally accepted for species delineation (Table S2e).
176 However, other OGRIs unanimously indicated that strains from this sub-clade belong to the
177 same species (Table S2a-d). A description of the novel species corresponding to sub-clade B,
178 for which the name *Allorhizobium ampelinum* sp. nov. is proposed, is given below.

179 The third genomic species comprised strain Av2 alone (Figs. 1 and S1, Table S2). To get a
180 more comprehensive insight into the diversity of the *All. vitis* species complex, we conducted
181 a second phylogenomic analysis where we included 34 additional genomes of *All. vitis* that
182 were available in GenBank but not yet published (Table S1b). Based on core-genome
183 phylogeny and ANIb calculations (Fig. S2, Table S3), additional strains were taxonomically
184 assigned as *All. vitis sensu stricto* (sub-clade A) and *All. ampelinum* (sub-clade B). Strain Av2
185 then grouped with three other strains originating from the USA (sub-clade D; Fig. S2). These
186 four strains comprised in the sub-clade D were genetically very similar and exhibited >99.8
187 ANI between each other (Table S3). Moreover, additional sub-clades C and E were apparent,

188 corresponding to two other new genomic species of *All. vitis* species complex (Fig. S2, Table
189 S3). Genomic species corresponding to sub-clade C and sub-clade D were closely related, as
190 their ANIb values were in the range 94.62-94.93 %, which is slightly below the threshold for
191 species delimitation (~95-96%) (29).
192
193 **Pan-genome analyses**
194 A ML pan-genome phylogeny of the 64 *Rhizobiaceae* genome dataset was estimated from a
195 matrix of the presence or absence of 33,396 orthologous gene clusters (Fig. 2; Fig. S3). The
196 pan-genome phylogeny (Fig. 2; Fig. S3) presented the same resolved sub-clades of the *All.*
197 *vitis* complex as the core-genome phylogeny (Fig. 1). Furthermore, *Rhizobiaceae* genera and
198 clades were generally differentiated based on the pan-genome tree (Fig. 2; Fig. S3).
199 Nevertheless, some inconsistencies were observed: tumorigenic strain *Neorhizobium* sp.
200 NCHU2750 was more closely related to the representatives of the genus *Agrobacterium*,
201 while nodulating *Pararhizobium giardinii* H152^{T} was grouped with *Ensifer* spp. (Fig. 2; Fig.
202 S3). These inconsistencies were also observed in another pan-genome phylogeny inferred
203 using parsimony (data not shown). Such limitations of gene content-based phylogenies have
204 previously been reported (30, 31).
205 Focusing on 14 *All. vitis* species complex strains, we identified 10,501 pan-genome gene
206 clusters. The core-genome ('strict core' and 'soft core' compartments) of the species
207 complex comprised 3,775 gene clusters (35.95% of total gene clusters), with 3,548 gene
208 clusters strictly present in all 14 strains (Fig. 3). The accessory genome contained 4,516 in the
209 cloud (43% of total gene clusters) and 2,210 gene clusters in the shell (21.05% of total gene
210 clusters) (Fig. 3).
211

212 **Clade-specific gene clusters**

213 Homologous gene families specific to particular clades of interest, i.e. with contrasted
214 presence pattern with respect to closely related clades, were identified using both
215 Pantagruel or GET_HOMOLOGUES software packages. Both sets of inferred clade-specific
216 genes were to a large extent congruent, although some differences were observed (Table
217 S4), owing to the distinct approaches employed by these software packages (32, 33). We
218 focused on clusters of contiguous clade-specific genes for which we could predict putative
219 molecular functions or association to a biological process. The results are summarized below
220 and in Table S4.

221

222 *All. vitis* species complex

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

232 Half of the *Av*SC-specific genes are gathered in contiguous clusters for most of which we
233 could predict putative function (Table S4); most of the other half are scattered on
234 chromosome 1 and have unknown function. Predicted functions of clustered genes revealed
235 that they are strikingly convergent: most are involved in either environmental signal

236 perception (4 clusters), stress response (2 clusters), aromatic compound and secondary
237 metabolite biosynthesis (3 clusters) and/or aromatic compound degradation response (2
238 clusters). In addition, one cluster encodes a multicomponent K^+:H^+ antiporter, which is likely
239 useful for adaptation to pH changes, and three clusters harbor several ABC transporter
240 systems for sugar or nucleotide uptake. Finally, one cluster on chromosome 1 encodes a
241 putative auto-transporter adhesin protein, which may have a role in plant commensalism
242 and pathogenesis.

243 All studied *All. vitis* species complex strains carried a *pehA* gene encoding a
244 polygalacturonase enzyme. Unlike other agrobacteria, *All. vitis* strains are known to produce
245 a polygalacturonase, regardless of their tumorigenicity (34). However, this gene was present
246 also in *All. taibaishanense* 14971^T, *All. terrae* CC-HIH110^T and *All. oryziradicis* N19^T, but
247 absent in *All. undicola* ORS 992^T and in other studied members of the *Rhizobiaceae* family.

248 Furthermore, we detected the presence of gene encoding enzyme 1-aminocyclopropane-
249 1-carboxylate deaminase (*acdS*) in all studied *All. vitis* species complex strains. This gene is
250 considered to be important for plant-bacteria interaction through its involvement in
251 lowering the level of ethylene produced by the plant (35). We found this gene in all other
252 *Allorhizobium* spp., and in some other *Rhizobiaceae* (data not shown), including *R.*
253 *rhizogenes* strains. However, *acdS* gene was not present in *Agrobacterium* spp., even when
254 the similarity search (blastp) was extended to *Agrobacterium* spp. strains available in
255 GenBank, consistent with previous findings (36).

256 Tartrate utilization ability was previously reported for most of the *All. vitis* strains (37-39).
257 Therefore, we searched *All. vitis* species complex genomes for the presence of tartrate
258 utilization (TAR) regions. All strains except IPV-BO 6186 and IPV-BO 7105 carried TAR gene
259 clusters. Moreover, we could not find any *All. vitis*-like TAR regions in any other *Rhizobiaceae*

260 strain. Sequence comparison of TAR regions from *All. vitis* species complex strains using ANIb
261 algorithm (Table S5) showed they could be divided into four types (Fig. S4). The first type is
262 represented by a previously characterized TAR region called TAR-I, carried on the tartrate
263 utilization plasmid pTrAB3 of strain AB3 (39, 40). The second type included representatives
264 of TAR-II (carried on pTiAB3) and TAR-III (carried on pTrAB4) regions, which were previously
265 described to be related to each other (39, 41). A third TAR region type, which we designate
266 TAR-IV, was characterized by the absence of a second copy of *ttuC* gene (tartrate
267 dehydrogenase). The TAR-IV region type is found in *All. ampelinum* strain S4^T, in which the
268 tartrate utilization system is located on the large plasmid pAtS4c (initially named pTrS4) (40).
269 The TAR system of *Allorhizobium* sp. strain Av2 is a unique type (TAR-V), which is related to
270 region type TAR-I, but is characterized by the absence of the *ttuA* gene (a LysR-like
271 regulator). We compared the distribution of these TAR region types in strain genomes,
272 showing there is no TAR region type associated to any genomic species (Table S6). *All. vitis*
273 *sensu stricto* strains K309^T and KFB 253 carry a TAR-II/III region. In addition to TAR-II/III
274 region, strain KFB 239 carries a TAR-I region (Table S6), a combination similar to that found
275 in the well-characterized strain AB3 (39). *All. ampelinum* strains S4^T, IPV-BO 1861-5, KFB 264
276 and V80/94 contain a TAR-IV region, while the remaining *All. ampelinum* strains IPV-BO
277 5159, KFB 243, KFB 250 and KFB 254 additionally carry a TAR-II/III region (Table S6).
278
279 *All. vitis sensu stricto*
280 Using Pantagruel and GET_HOMOLOGUES pipelines, we identified 63 and 78 genes, that are
281 specific to *All. vitis sensu stricto* (Av-specific, present in all five strains and in none of *All.*
282 *ampelinum*), respectively. 32 of these *Av*-specific genes are clustered into four main loci in
283 the genome of strain K309^T, for which we could predict putative function (Table S4). One *Av*-

284 specific gene cluster (Av-GC1, Table S4) comprised genes functionally annotated to be
285 involved in the degradation process of salicylic acid and gentisic acid (2,5-dihydroxybenzoic
286 acid) (MetaCyc pathways PWY-6640 and PWY-6223). Av-GC1 was located on Contig 1
287 (LMVL02000001.1) of reference strain K309^T genome, which is likely part of the chromid,
288 based on its high ANI with the chromid (Chromosome 2) of strain S4^T, whose genome
289 sequence is complete. BLAST searches showed that this gene cluster is also present in some
290 representatives of *Agrobacterium deltaense*, i.e. *Agrobacterium* genomospecies G7 (data not
291 shown). Av-GC1 is predicted to encode the degradation of salicyl-CoA, an intermediate in
292 degradation of salicylic acid, to 3-fumarylpyruvate, via gentisic acid. Interestingly, strains KFB
293 239, IPV-BO 6186 and IPV-BO 7105 carried additional genes encoding the degradation of
294 salicylaldehyde to salicyl-CoA via salicylic acid and salicyl adenylate, as well as the gene
295 encoding the final step of gentisic acid degradation, the conversion of 3-fumarylpyruvate to
296 fumarate and pyruvate. The three strains encoding enzymes of the complete pathway for
297 degradation of salicylic acid and gentisic acid, and remaining strains K309^T and KFB 253
298 carrying a partial gene cluster, were phylogenetically separated and formed distinct sub-
299 clades within *All.* vitis *sensu stricto* (Fig. 1).

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

305 In addition, *Av*-specific gene clusters Av-GC2 and Av-GC3 (Table S4) were both predicted
306 to be involved in amino-acid uptake and catabolism. However, we were not able to predict
307 the precise molecular function of proteins and substrates of enzymes encoded by these *Av*-

308 specific gene clusters. Both these gene clusters are likely located on a putative plasmid, as
309 suggested by the presence of plasmid-related genes (replication and/or conjugation
310 associated genes) on the same contigs.
311
312 *All. ampelinum*
313 Based on Pantagruel and GET_HOMOLOGUES analyses, we identified 97 genes and 128
314 genes, respectively, that are specific to *All. ampelinum* (*Aa*-specific, present in all 8 strains
315 and in none of *All. vitis sensu stricto*). Taking advantage of the finished status of strain $S4^T$
316 genome, we found that 52/97 specific genes identified by Pantagruel occur on plasmids
317 rather than chromosomes. This is a significant over-representation compared to the
318 distribution of all genes (21.4% on plasmids, Chi-squared test *p*-value < 10^{-6}) or core-genome
319 genes (5.8% on plasmids, Chi-squared test *p*-value < 10^{-16}). For eleven contiguous gene
320 clusters we could predict putative function (Table S4). The *Aa*-specific gene clusters encode a
321 variety of putative biological functions; an enrichment analysis of their functional
322 annotations revealed a set of high-level biological processes that were over-represented:
323 transport and metabolism of amino-acids or polyamines like putrescine (three separate
324 clusters), lysin biosynthesis (two separate clusters), and nickel assimilation. The latter
325 function is predicted for gene cluster Aa-GC10, which is located on the 631-kb megaplasmid
326 pAtS4e and encodes the NikABCDE Ni^{2+} import system and a nickel-responsive
327 transcriptional regulator NikR. Aa-GC10 additionally includes genes with predicted functions
328 such as cation-binding proteins and a chaperone/thioredoxin, which may be involved in the
329 biosynthesis of ion-associated cofactors.
330

331 **Phenotypic and MALDI-TOF MS characterization**

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

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

350 We also tested the ability of *All. vitis* species complex strains to metabolize L-tartaric acid
351 and produce alkali from this compound. In the present study, we included only strains that
352 were not tested in our former work (42). Taken together, all tested *All. vitis* species complex
353 strains (Table 1) were able to produce alkali from tartrate. Interestingly, strains IPV-BO 6186

354 and IPV-BO 7105, for which we could not identify TAR gene clusters, were also positive for
355 this test.

356 As a broader way to characterize and phenotypically distinguish strains, we used MALDI-
357 TOF mass-spectrometry (MS) of pure bacterial cultures. MALDI-TOF MS revealed diversity
358 among the tested strains, while allowing to discriminate genomic species (Fig. S5).

359

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

361 As indicated by the core-genome phylogeny, the genus *Allorhizobium* is clearly separated
362 from the other representatives of the family *Rhizobiaceae,* including the "*R. aggregatum*
363 complex", which, with the genus *Ciceribacter*, formed a well-delineated sister clade to
364 *Allorhizobium* clade (Figs. 1, S1 and S2). The genome-based comparisons showed a clear
365 divergence between these two clades. In particular, members of the genus *Allorhizobium*
366 shared >74.9% AAI among each other, and 70.79-72.63% AAI with members of the "*R.*
367 *aggregatum* complex"/*Ciceribacter* clade (Table S7). On the other hand, representatives of
368 the genera *Shinella*, *Ensifer* and *Pararhizobium* showed 71.46-75.85% AAI similarity between
369 genera. Similarly, representatives of genera *Neorhizobium* and *Pseudorhizobium* showed
370 72.24-76.18 % AAI similarity between genera. In other words, AAI values suggested that the
371 existing genera *Ensifer*, *Pararhizobium* and *Shinella*, or *Neorhizobium* and *Pseudorhizobium*
372 were more closely related than the genus *Allorhizobium* and the "*R. aggregatum*
373 complex"/*Ciceribacter* clade. gANI and POCP values similarly supported the divergence of
374 the members of *Allorhizobium* genus and the "*R. aggregatum* complex"/*Ciceribacter* clade
375 (Table S7). Members of the genus *Allorhizobium* exhibited gANI and POCP values ranging
376 73.55-76.86 and 55.27-66.17, respectively, when compared with members of the "*R.*
377 *aggregatum* complex"/*Ciceribacter* clade, values that were similar to these seen between

378 representatives of the genera *Agrobacterium* and *Neorhizobium* (gANI 74.66-77.45; POCP
379 59.96-65.58).
380
381 **Discussion**
382 ***Allorhizobium vitis* is not a single species**
383 Genomic analyses allowed us to unravel the substantial taxonomic diversity within *All. vitis*.
384 In particular, whole-genome sequence comparisons and phylogenomic analyses clearly
385 showed that *Allorhizobium vitis* is not a single species, but represents a species complex
386 composed of several genomic species. Similarly, *Agrobacterium* biovar 1 (i.e. *A. tumefaciens*)
387 was initially considered a single species, but was later designated as a species complex
388 comprising closely related, but distinct genomic species. Several studies applying DDH
389 initially demonstrated this species diversity within *Agrobacterium* biovar 1 (43-45), which
390 was later supported by results obtained with AFLP (46, 47), housekeeping gene analysis (10,
391 11, 14) and whole-genome sequence analysis (30). Although Ophel and Kerr (9) also
392 performed DDH for several *All. vitis* strains, diversity within this species remained unknown
393 because these authors only studied strains that belonged to *All. vitis sensu stricto* as defined
394 here.
395 Our previous study based on the analysis of several housekeeping gene sequences
396 suggested the existence of several phylogenetic groups within *All. vitis* species complex (28).
397 The present study focused on two phylogenetic groups defined in our previous study: the
398 first comprises the type strain of *All. vitis* (strain K309^T) (9, 48), whereas the second includes
399 the well-characterized and completely sequenced strain S4^T (18). Consequently, we
400 amended the description of *All. vitis*, which now refers to the limited group within *All. vitis*

401 species complex strains (*All. vitis sensu stricto*) and proposed a description of a novel
402 species, *All. ampelinum* sp. nov. (see formal description below).

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

412

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

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

419 We also searched genomes of *All. vitis* species complex strains for genes and gene
420 clusters that were previously reported as important for the ecology of this bacterium. In this
421 regard, polygalacturonase production, a trait associated with grapevine root necrosis (34,
422 49, 50), and tartrate degradation (37) were proposed to contribute to the specialization of
423 *All. vitis* to its grapevine host. In addition, polygalacturonase activity might be involved in the
424 process of the invasion of the host plant, as postulated previously for other rhizobia (51).

425 Although all *All. vitis* species complex strains carried the *pehA* gene encoding a
426 polygalacturonase enzyme, this gene was not restricted to this bacterial group, as it was also
427 present in all other *Allorhizobium* spp. strains included in our analysis, except for *All.*
428 *undicola*.

429 All *All. vitis* species complex strains included in this study, except for strains IPV-BO 6186
430 and IPV-BO 7105, carried TAR regions. However, all of them were able to metabolize tartrate
431 and produce alkali from this compound. Therefore, we speculate that strains IPV-BO 6186
432 and IPV-BO 7105 must carry another type of TAR system, distinct from those described so far
433 in other *All. vitis* strains. Furthermore, some diversity between TAR regions and variable
434 distribution patterns of different TAR regions among strains were observed, in line with
435 previously reported data (39). The existence of non-tartrate-utilizing strains was also
436 documented in the literature (39). Considering the fact that tartrate utilization in *All. vitis*
437 has only been observed as plasmid-borne (40, 41, 52), this suggests that tartrate utilization is
438 an accessory trait that can be readily gained via the acquisition of a plasmid encoding this
439 trait and selected for in tartrate-abundant environments. Because grapevine is rich in
440 tartrate (53), utilization of this substrate may enhance the competitiveness of *All. vitis*
441 species complex strains in colonizing this plant species (37).

442 We observed that an important fraction of the species-specific genes for *All. vitis sensu*
443 *stricto* and *All. ampelinum* occurred on chromids and plasmids, suggesting that these
444 replicons may be an important part of these species' adaptive core-genome, as previously
445 observed in the *A. tumefaciens* species complex (30). Ecological differentiation of the two
446 main species of the *All. vitis* species complex seems to rely on consumption of different
447 nutrient sources, including polyamines and nickel ion (potentially as a key cofactor of

448 ecologically important enzymes) for *All. ampelinum*, and phenolic compounds for *All. vitis*
449 *sensu stricto*.

450 Even though *All. vitis sensu stricto* strains carried a putative gene cluster which predicted
451 function was the degradation of gentisic acid, we could not experimentally demonstrate this
452 trait. Gentisic acid was detected in grapevine leaves (54) and is likely present in other parts
453 of this plant. This compound was reported as a plant defense signal that can accumulate in
454 some vegetable plants responding to compatible viral pathogens (55, 56). In addition, a sub-
455 clade within *All. vitis sensu stricto* composed of strains K309^T and KFB 253 carried a complete
456 pathway for degradation of salicylic acid through gentisic acid. Salicylic acid is recognized as
457 an important molecule for plant defense against certain pathogens (57). The role of salicylic
458 and gentisic acid in grapevine defense mechanism against pathogenic bacteria has not been
459 studied in detail, and further investigations are required to understand their effect against
460 tumorigenic agrobacteria. Furthermore, we predicted, and demonstrated that all studied *All.*
461 *vitis sensu stricto* strains have the specific ability to degrade 4-hydroxyphenylacetate, an
462 activity that may contribute to the detoxication of aromatic compounds and thus to the
463 survival of this bacterium in soil, notably in competition against bacteria lacking this
464 pathway.

465 Similarly, gene clusters putatively involved in polyamine metabolism or nickel assimilation
466 might confer to *All. ampelinum* the ability to persist in harsh environments. In this respect,
467 nickel import has been shown to be essential for hydrogenase function in *E. coli* (58).
468 Hydrogenase function has in turn been proposed as a potential mechanism for detoxication
469 of phenolic compounds in *A. vitis* (59) and may thus have an important role in survival in the
470 rhizosphere.

471

472 **Delineation of the genus *Allorhizobium***

473 The genus *Allorhizobium* was clearly differentiated from other *Rhizobiaceae* genera based on
474 core- and pan-genome-based phylogenies, in line with previous studies employing genome-
475 wide phylogeny (15, 16). We included diverse *All. vitis* species complex strains into our
476 analysis, confirming that these bacteria, principally recognized as grapevine crown gall
477 causative agents, belong to the genus *Allorhizobium*.

478 On the other hand, the taxonomic status of the "*R. aggregatum* complex"/*Ciceribacter*
479 clade is still unresolved. Although MLSA suggested that "*R. aggregatum* complex" is a sister
480 clade of the genus *Agrobacterium*, the more thorough phylogenetic analyses performed in
481 this study rather showed that the "*R. aggregatum* complex" grouped with *Ciceribacter* spp.,
482 in a clade that is more closely related to the genus *Allorhizobium*. Presently, there is no
483 widely accepted criteria and scientific consensus regarding the delineation of new bacterial
484 genera (27). In this study, existing *Rhizobiaceae* genera were compared using several
485 delineation methods proposed in the literature, such as AAI (60, 61), POCP (62) and gANI/AF
486 (63), which we complemented with genome-based phylogenies. Taken together, our
487 genome-based analysis suggested that *Allorhizobium* represents a genus clearly separated
488 from other *Rhizobiaceae* genera, including closely related "*R. aggregatum*
489 complex"/*Ciceribacter* clade. A separate and more focused analysis is however required to
490 explore the taxonomic diversity and structure of the "*R. aggregatum* complex"/*Ciceribacter*
491 clade.

492

493 **Conclusions**

494 Whole-genome sequence comparisons and phylogenomic analyses placed *All. vitis* strains
495 within the genus *Allorhizobium*, which was clearly differentiated from other *Rhizobiaceae*

496 genera, including the closely related "*R. aggregatum* complex"/*Ciceribacter* clade. We
497 revealed an extensive and structured genomic diversity within *All. vitis*, which in fact
498 represents a species complex composed of several genomic species. Consequently, we
499 emended the description of *All. vitis*, now encompassing a restricted group of strains within
500 the *All. vitis* species complex (i.e. *All. vitis sensu stricto*) and proposed a description of a
501 novel species, *All. ampelinum* sp. nov. Further analyses including pan-genome reconstruction
502 and phylogeny-driven comparative genomics revealed loci of genomic differentiation
503 between these two species. Functional analysis of these species-specific loci suggested that
504 these species are ecologically differentiated as they can consume specific nutrient sources
505 (*All. ampelinum*), or degrade specific toxic compounds (*All. vitis sensu stricto*). We identified
506 another two potential genomic species within the *All. vitis* species complex, for which the
507 limited diversity of available isolates prevented further characterization. We also described
508 how accessory genomic regions associated with the colonization of grapevine host plant are
509 distributed across species, and how they combine to form diverse genotypes. However,
510 given the complete bias in sampling of *All. vitis* strains – all grapevine pathogens – the
511 ecological significance of this genetic diversity remains unclear. We encourage future studies
512 to integrate genomic data from new genomically diverse isolates, to further unravel the
513 ecological basis of *All. vitis* species complex diversification.
514

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

517 The description of *Agrobacterium vitis* is provided by Ophel and Kerr (9). Young et al. (13)
518 proposed the transfer of *A. vitis* to the genus *Rhizobium*, but it was neither widely accepted
519 by the scientific community nor supported by further studies (14, 64, 65). Mousavi et al. (11)

520 reclassified this species to the genus *Allorhizobium*, which was included into the Validation
521 list no. 172 of the IJSEM (66). Hördt et al. (16) emended a description of *All. vitis* by including
522 genome sequence data for its type strain, which was published in the List of changes in
523 taxonomic opinion no. 32 (67).

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

534 The whole-genome sequence of type strain K309^T is available in GenBank under the
535 accessions LMVL00000000.2 and GCA_001541345.2 for the Nucleotide and Assembly
536 databases, respectively (48). The genomic G+C content of the type strain is 57.55%. Its
537 approximate genome size is 5.75 Mbp.

538 Basonym: *Agrobacterium vitis* Ophel and Kerr 1990.

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

542

543 **Description of *Allorhizobium ampelinum* sp. nov.**

544 The description and properties of the new species are given in the protologue (Table 2).

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

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

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

554

555 **Table 2**

556

557 **Methods**

558 ***Allorhizobium vitis* strains**

559 *All. vitis* strains used in this study were isolated from crown gall tumors on grapevine
560 originating from different geographical areas (Table 1). These strains were predominantly
561 representatives of the two main phylogenetic groups (C and D) delineated in our previous
562 study (28).

563

564 **DNA extraction**

565 For whole genome sequencing, genomic DNA was extracted from bacterial strains grown on
566 King's medium B (King et al. 1954) at 28°C for 24 h using NucleoSpin Microbial DNA kit

567 (Macherey-Nagel, Germany). The quality of the genomic DNA was assessed by
568 electrophoresis in 0.8% agarose gel.
569
570 **Genome sequencing**
571 Draft whole-genome sequences were obtained for 11 *All. vitis* strains (Table 1). DNA libraries
572 were obtained with Nextera XT DNA Library Prep Kit (Illumina, USA). Paired-end sequencing
573 (2 × 300 bp) was performed on an Illumina MiSeq platform generating 487,883- 2,309,377
574 paired reads per genome. Trimming and quality filtering of raw reads were conducted using
575 Trimmomatic (Galaxy Version 0.36.5) (68) implemented on the Galaxy Web server (69). The
576 read quality was assessed with FastQC (Galaxy Version 0.72+galaxy1)
577 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). In order to achieve higher
578 coverage for strains Av2, IPV-BO 1861-5, KFB239 and KFB 264, additional paired-end
579 sequencing (2 × 150 bp) was performed using an Illumina NextSeq 500 platform generating
580 1,037,619-1,443,575 paired reads. Demultiplexing and adapter clipping was done using the
581 bcl2fastq(2) conversion software (Illumina, USA).
582
583 **Genome assembly and annotation**
584 *De novo* genome assemblies were performed using the SPAdes genome assembler (Galaxy
585 Version 3.12.0+galaxy1) (70). For genomes sequenced on the MiSeq and NextSeq platforms,
586 both sets of reads were used for assembly. The genome sequences were deposited to
587 DDBJ/ENA/GenBank under the Whole Genome Shotgun projects accession numbers listed in
588 Table 1, under BioProject ID PRJNA557463.
589 The genome sequences were annotated using Prokka (Galaxy Version 1.13) (71) and NCBI
590 Prokaryotic Genomes Annotation Pipeline (PGAP) (72). Prokka Version 1.14.6 was used to

591 annotate genomes as a part of the Pantagruel pipeline (task 0; see below and
592 Supplementary Methods). Functional annotation of proteins encoded by each gene family
593 clustered by Pantagruel was conducted by the InterProScan software package Version 5.42-
594 78.0 (Jones et al., 2014) as implemented in the Pantagruel pipeline (Task 4). Additionally,
595 annotation of particular sequences of interest and metabolic pathway prediction were
596 performed using BlastKOALA and GhostKOALA (last accessed on December, 2020) (73).
597 Protein sequences analyzed were subjected to Pfam domain searches (database release
598 32.0, September 2018, 17929 entries) (74). Metabolic pathway prediction was performed
599 using KEGG (75) and MetaCyc (76) databases (last accessed on December, 2020).
600 The NCBI BLASTN and BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi), as well as
601 BLAST search tool of KEGG database (last accessed on December, 2020) (75), were used for
602 *ad-hoc* sequence comparisons at the nucleotide and amino acid levels, respectively.
603

604 **Core- and pan-genome phylogenomic analyses**

605 For phylogenomic analysis, whole genome sequences of 69 *Rhizobiaceae* strains were used,
606 including 14 strains of *All. vitis* (Table 1) and 55 reference *Rhizobiaceae* strains (Table S1a).
607 Additionally, in order to further explore the phylogenetic diversity of *All. vitis*, another core-
608 genome phylogeny was inferred from an extended dataset that also included 34 *All. vitis*
609 genomes available from GenBank but not yet published in peer-review journals by sequence
610 depositors (Table S1b). To build phylogenies based on the core-genome (supermatrix of
611 concatenated non-recombining core gene alignments) and on the pan-genome (homologous
612 gene cluster presence/absence matrix), we used the GET_HOMOLOGUES Version 10032020
613 (33) and GET_PHYLOMARKERS Version 2.2.8.1_16Jul2019 (77) software packages. Details of
614 the bioinformatic pipeline and used options are described in the Supplementary Methods.

615

616 **Overall genome relatedness indices**

617 To differentiate between the strains, different OGRIs were computed. For species
618 delimitation, we relied on the values of average nucleotide identity (ANI) (29, 78) and digital
619 DNA-DNA hybridization (dDDH) (79) among strain genomes. Because different
620 implementations of the ANI metric are known to give slightly different results (80), ANI was
621 calculated using several programs: PyANI Version 0.2.9 (for metrics ANIb and ANIm) (81)
622 (https://github.com/widdowquinn/pyani), OrthoANIu Version 1.2 (82) and FastANI Version
623 1.2 (83) tools. dDDH values were calculated using the Genome-to-Genome Distance
624 Calculator (GGDC) Version 2.1 (79).

625 For genus delimitation, we relied on average amino acid identity (AAI) (22, 60, 78),
626 genome-wide average nucleotide identity (gANI) and alignment fraction (AF) (84), and the
627 percentage of conserved proteins (POCP) (62). AAI values were calculated with CompareM
628 Version 0.0.23 (https://github.com/dparks1134/CompareM). gANI and AF values were
629 obtained by the ANIcalculator Version 1.0 (84). POCP values were calculated using
630 GET_HOMOLOGUES software package (33). Details of the used software and options are
631 given in the Supplementary Methods.

632

633 **Genome gene content analyses and identification of clade-specific genes**

634 To explore the distribution of genome gene contents, we conducted further pan-genome
635 analyses on more focused datasets, using two different bioinformatics pipelines, from which
636 we present a consensus. Firstly, a pan-genome database was constructed using the
637 Pantagruel pipeline Version 00aaac71f85a2afa164949b86fbc5b1613556f36 under the
638 default settings as described previously (31, 32) and in Supplementary Methods. Because of

639 computationally intensive tasks undertaken in this pipeline, the dataset was limited to the
640 *Allorhizobium* genus and its sister clade "*Rhizobium aggregatum* complex"/*Ciceribacter* (28
641 strains).

642 Secondly, we analyzed a more focused dataset comprised of the 14 *All. vitis* species
643 complex strains (Table 1) and four *Allorhizobium* spp. (*All. oryziradicis* N19^T, *All.*
644 *taibaishanense* 14971^T, *All. terrae* CC-HIH110^T and *All. undicola* ORS 992^T; Table S1a), using
645 the GET_HOMOLOGUES software package (33). Pan-genome gene clusters were classified
646 into core, soft core, cloud and shell compartments (85) and species-specific gene families
647 were identified from the pan-genome matrix. For details on the used scripts and options, see
648 Supplementary Methods.

649

650 **Biochemical tests**

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

655 The metabolism of 4-hydroxyphenylacetic acid (p-hydroxyphenylacetic acid; Acros
656 Organics, Product code: 121710250) and gentisic acid (2,5-dihydroxybenzoic acid; Merck,
657 Product Number: 841745) was performed in AT minimal medium (86, 87) supplemented
658 with yeast extract (0.1 g/L), bromthymol blue (2.5 ml/L of 1 % [w/v] solution made in 50%
659 ethanol), and the tested compound (1 g/L). Hydroxyphenylacetic and gentisic acid were
660 added as filter-sterilized 1% aqueous solutions. Bacterial growth and color change of the
661 medium were monitored during one week of incubation at 28°C and constant shaking (200

662 rpm/min). Metabolism of L(+)-tartaric acid, involving production of alkali from this
663 compound, was tested as described before (5).
664

665 **MALDI-TOF Mass Spectrometry analysis**

666 Sample preparation for MALDI-TOF mass spectrometry (MS) was carried out according to the
667 Protocol 3 described by Schumann and Maier (88). Instrument settings for the
668 measurements were as described previously by Tóth et al. (89). The dendrogram was
669 created using the MALDI Biotyper Compass Explorer software (Bruker, Version 4.1.90).
670

671 **Abbreviations**

672 AAI: Average amino acid identity; AF: Alignment fraction; ANI: Average nucleotide identity;
673 DDH: DNA-DNA hybridization; dDDH: Digital DDH; gANI: Genome-wide ANI; MS: Mass-
674 spectrometry; OGRI: Overall genome relatedness index; MLSA: Multi-locus sequence
675 analysis; PGAP: Prokaryotic genomes annotation pipeline; POCP: Percentage of conserved
676 proteins; TAR: Tartrate utilization; Ti: Tumor-inducing.
677

678 **Acknowledgments**

679 We would like to thank Cathrin Spröer and Boyke Bunk for conducting the Illumina
680 sequencing. We are grateful to Anja Frühling and Ulrike Steiner for support in phenotypic
681 tests.
682

683 **Authors' contributions**

684 NK and FL conceived and designed the study, and analyzed data. FL conceived and
685 implemented the bioinformatic analysis pipeline Pantagruel. SV performed phenotypic

686 characterization of the strains (API, Biolog and MALDI-TOF). EB provided bacterial strains. JP,
687 JO and KS were involved in interpreting data. NK and FL wrote the manuscript. All authors
688 read, discussed, edited and approved the final manuscript.
689

690 **Funding**

691 This research was supported by the Georg Forster Fellowship for postdoctoral research from
692 the Alexander von Humboldt-Foundation, Bonn, Germany. The work of FL was supported by
693 by a Medical Research Council (MRC) grant (MR/N010760/1) and the Wellcome Trust Grant
694 [206194]. Bioinformatic analyses were supported by the BMBF-funded de.NBI Cloud within
695 the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A,
696 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).
697

698 **Availability of data and materials**

699 The genome sequences generated in this study were deposited to DDBJ/ENA/GenBank
700 under the Whole Genome Shotgun projects accession numbers listed in Table 1, under
701 BioProject ID PRJNA557463. The versions described in this paper are first versions.
702 All other relevant data (including output of analyses) referring to this project have been
703 deposited on Figshare under the project accession 20894, available at figshare
704 (https://figshare.com/), with individual items accessible at DOIs:
705 10.6084/m9.figshare.17105267, 10.6084/m9.figshare.17125571,
706 10.6084/m9.figshare.16850071, 10.6084/m9.figshare.16849165,
707 10.6084/m9.figshare.13440218 and 10.6084/m9.figshare.17125568.
708 (DOIs to be published upon manuscript acceptance; private links for peer-review:
709 https://figshare.com/s/b6c7ccc50f0179a08ce6,

710 https://figshare.com/s/6aaae45f5eb5a81baeb5,

711 https://figshare.com/s/c88e7412cf58bc9b1b54,

712 https://figshare.com/s/77bb93622b1bf788bada,

713 https://figshare.com/s/f17bb1c00a960714e097,

714 https://figshare.com/s/eecc49ea166bbe061554).

715

716 **Declarations**

717

718 **Ethics approval and consent to participate**

719 Not applicable.

720

721 **Consent for publication**

722 Not applicable.

723

724 **Competing interests**

725 The authors declare no competing interests.

726

727 **Author details**

728 [1]Julius Kühn Institute, Federal Research Centre for Cultivated Plants (JKI), Institute for

729 Epidemiology and Pathogen Diagnostics, Messeweg 11-12, 38104 Braunschweig, Germany.

730 [2]Julius Kühn Institute, Federal Research Centre for Cultivated Plants (JKI), Institute for Plant

731 Protection in Horticulture and Forests, Messeweg 11-12, 38104 Braunschweig, Germany

732 (current affiliation). [3]Alma Mater Studiorum - University of Bologna, Viale G. Fanin, 42,

733 40127 Bologna, Italy. [4]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell

Cultures, Inhoffenstrasse 7B, 38124 Braunschweig, Germany. [5]Research Institute of Horticulture, ul. Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland. [6]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. [7]Wellcome Sanger Institute, Pathogens and Microbes Programme, Wellcome Genome Campus, Hinxton, Saffron Walden, CB10 1RQ, United Kingdom

References

1. Kuzmanović N, Pulawska J, Hao L, Burr TJ. The ecology of *Agrobacterium vitis* and management of crown gall disease in vineyards. Curr Top Microbiol Immunol. 2018.
2. Sawada H, Ieki 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, 359[AL]. 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 Bacteriol. 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;44(2):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, Osterman J, Wahlberg 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 undicola* 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 undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium*

775 *radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. Int J Syst Evol Microbiol. 2001;51(Pt 1):89-
776 103.
777 14.Costechareyre D, Rhouma A, Lavire C, Portier P, Chapulliot D, Bertolla F, et al. Rapid and efficient
778 identification of *Agrobacterium* species by *recA* allele analysis: *Agrobacterium recA* diversity. Microb
779 Ecol. 2010;60(4):862-72.
780 15.Ormeño-Orrillo E, Servín-Garcidueñas LE, Rogel MA, González V, Peralta H, Mora J, et al.
781 Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. Syst Appl
782 Microbiol. 2015;38(4):287-91.
783 16.Hördt A, López MG, Meier-Kolthoff JP, Schleuning M, Weinhold L-M, Tindall BJ, et al. Analysis of
784 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria.
785 Front Microbiol. 2020;11(468).
786 17.Ramírez-Bahena MH, Vial L, Lassalle F, Diel B, Chapulliot D, Daubin V, et al. Single acquisition of
787 protelomerase gave rise to speciation of a large and diverse clade within the
788 *Agrobacterium*/*Rhizobium* supercluster characterized by the presence of a linear chromid. Mol
789 Phylogenet Evol. 2014;73:202-7.
790 18.Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, et al. Genome sequences of
791 three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria.
792 J Bacteriol. 2009;191(8):2501-11.
793 19.Harrison PW, Lower RP, Kim NK, Young JP. Introducing the bacterial 'chromid': not a chromosome,
794 not a plasmid. Trends Microbiol. 2010;18(4):141-8.
795 20.Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allardet-Servent A. Unconventional
796 genomic organization in the alpha subgroup of the Proteobacteria. J Bacteriol. 1998;180(10):2749-55.
797 21.Tanaka K, Urbanczyk H, Matsui H, Sawada H, Suzuki K. Construction of physical map and mapping
798 of chromosomal virulence genes of the biovar 3 *Agrobacterium* (*Rhizobium vitis*) strain K-Ag-1. Genes
799 Genet Syst. 2006;81(6):373-80.
800 22.Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol.
801 2005;187(18):6258-64.
802 23.Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the *Bacteria* and
803 *Archaea*. Int J Syst Evol Microbiol. 2014;64(2):316-24.
804 24.Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized
805 bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol.
806 2018;36(10):996-1004.
807 25.Wayne LG, Brenner DJ, Colwell RR, Grimont PaD, Kandler O, Krichevsky MI, et al. Report of the ad
808 hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol. 1987;37.
809 26.Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA
810 sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol. 1994;44.
811 27.de Lajudie PM, Andrews M, Ardley J, Eardly B, Jumas-Bilak E, Kuzmanovic N, et al. Minimal
812 standards for the description of new genera and species of rhizobia and agrobacteria. Int J Syst Evol
813 Microbiol. 2019;69(7):1852-63.
814 28.Kuzmanović N, Biondi E, Bertaccini A, Obradović A. Genetic relatedness and recombination
815 analysis of *Allorhizobium vitis* strains associated with grapevine crown gall outbreaks in Europe. J
816 Appl Microbiol. 2015;119:786-96.
817 29.Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species
818 definition. Proc Natl Acad Sci USA. 2009;106(45):19126-31.
819 30.Lassalle F, Planel R, Penel S, Chapulliot D, Barbe V, Dubost A, et al. Ancestral genome estimation
820 reveals the history of ecological diversification in *Agrobacterium*. Genome Biol Evol. 2017;9(12):3413-
821 31.
822 31.Lassalle F, Dastgheib SMM, Zhao F-J, Zhang J, Verbarg S, Frühling A, et al. Phylogenomics reveals
823 the basis of adaptation of *Pseudorhizobium* species to extreme environments and supports a
824 taxonomic revision of the genus. Syst Appl Microbiol. 2021;44(1):126-65.
825 32.Lassalle F, Veber P, Jauneikaite E, Didelot X. Automated reconstruction of all gene histories in
826 large bacterial pangenome datasets and search for co-evolved gene modules with Pantagruel.
827 bioRxiv. 2019:586495.

828 33.Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable
829 and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79(24):7696-701.
830 34.McGuire RG, Rodriguez-Palenzuela P, Collmer A, Burr TJ. Polygalacturonase production by
831 *Agrobacterium tumefaciens* biovar 3. Appl Environ Microbiol. 1991;57(3):660-4.
832 35.Gamalero E, Glick BR. Bacterial modulation of plant ethylene levels. Plant Physiol. 2015;169(1):13-
833 22.
834 36.Bruto M, Prigent-Combaret C, Muller D, Moënne-Loccoz Y. Analysis of genes contributing to plant-
835 beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria. Scientific
836 Reports. 2014;4(1):6261.
837 37.Salomone JY, Szegedi E, Cobanov P, Otten L. Tartrate utilization genes promote growth of
838 *Agrobacterium* spp. on grapevine. Mol Plant-Microbe Interact. 1998;11(8):836-8.
839 38.Szegedi E. Host range and specific L(+)tartrate utilization of biotype 3 of *Agrobacterium*
840 *tumefaciens*. Acta Phytopathol Acad Sci Hung. 1985;20:17-22.
841 39.Salomone JY, Crouzet P, De Ruffray P, Otten L. Characterization and distribution of tartrate
842 utilization genes in the grapevine pathogen *Agrobacterium vitis*. Molecular plant-microbe
843 interactions : MPMI. 1996;9(5):401-8.
844 40.Szegedi E, Otten L, Czakó M. Diverse types of tartrate plasmids in *Agrobacterium tumefaciens*
845 biotype III strains. Mol Plant-Microbe Interact. 1992;5:435-8.
846 41.Crouzet P, Otten L. Sequence and mutational analysis of a tartrate utilization operon from
847 Agrobacterium vitis. J Bacteriol. 1995;177(22):6518-26.
848 42.Kuzmanović N, Ivanović M, Prokić A, Gašić K, Zlatković N, Obradović A. Characterization and
849 phylogenetic diversity of *Agrobacterium vitis* from Serbia based on sequence analysis of 16S-23S
850 rRNA internal transcribed spacer (ITS) region. Eur J Plant Pathol. 2014;140:757-68.
851 43.De Ley J. Phylogeny of procaryotes. Taxon. 1974;23:291–300.
852 44.De Ley J, Tijtgat R, De Smedt J, Michiels M. Thermal stability of DNA: DNA hybrids within the
853 genus *Agrobacterium*. J Gen Microbiol. 1973;78(2):241-52.
854 45.Popoff MY, Kersters K, Kiredjian M, Miras I, Coynault C. Taxonomic position of *Agrobacterium*
855 strains of hospital origin. Annales de microbiologie. 1984;135a(3):427-42.
856 46.Portier P, Fischer-Le Saux M, Mougel C, Lerondelle C, Chapulliot D, Thioulouse J, et al.
857 Identification of Genomic Species in *Agrobacterium* Biovar 1 by AFLP Genomic Markers. Appl Environ
858 Microbiol. 2006;72(11):7123-31.
859 47.Mougel C, Thioulouse J, Perriere G, Nesme X. A mathematical method for determining genome
860 divergence and species delineation using AFLP. Int J Syst Evol Microbiol. 2002;52(Pt 2):573-86.
861 48.Gan HM, Lee MVJ, Savka MA. High-quality draft genome sequence of the type strain of
862 *Allorhizobium vitis*, the primary causal agent of grapevine crown gall. Microbiol Resour Announc.
863 2018;7(9):e01045-18.
864 49.Rodriguez-Palenzuela P, Burr TJ, Collmer A. Polygalacturonase is a virulence factor in
865 *Agrobacterium tumefaciens* biovar 3. J Bacteriol. 1991;173(20):6547-52.
866 50.Brisset M-N, Rodriguez-Palenzuela P, Burr TJ, Collmer A. Attachment, chemotaxis, and
867 multiplication of *Agrobacterium tumefaciens* biovar 1 and biovar 3 on grapevine and pea. Appl
868 Environ Microbiol. 1991;57(11):3178-82.
869 51.Muñoz JA, Coronado C, Pérez-Hormaeche J, Kondorosi A, Ratet P, Palomares AJ. MsPG3, a
870 *Medicago sativa* polygalacturonase gene expressed during the alfalfa-*Rhizobium meliloti* interaction.
871 Proc Natl Acad Sci USA. 1998;95(16):9687-92.
872 52.Otten L, Crouzet P, Salomone JY, de Ruffray P, Szegedi E. *Agrobacterium vitis* strain AB3 harbors
873 two independent tartrate utilization systems, one of which is encoded by the Ti plasmid. Mol Plant-
874 Microbe Interact. 1995;8:138-46.
875 53.Ruffner HP. Metabolism of tartaric and malic acids in *Vitis* : A review - Part A. Vitis: Journal of
876 Grapevine Research. 2016;21:247-.
877 54.Pantelić MM, Zagorac DČD, Ćirić IŽ, Pergal MV, Relić DJ, Todić SR, et al. Phenolic profiles,
878 antioxidant activity and minerals in leaves of different grapevine varieties grown in Serbia. J Food
879 Compos Anal. 2017;62:76-83.

880 55.Bellés JM, Garro R, Fayos J, Navarro P, Primo J, Conejero V. Gentisic acid as a pathogen-inducible
881 signal, additional to salicylic acid for activation of plant defenses in tomato. Molecular Plant-Microbe
882 Interactions®. 1999;12(3):227-35.
883 56.Bellés JM, Garro R, Pallás V, Fayos J, Rodrigo I, Conejero V. Accumulation of gentisic acid as
884 associated with systemic infections but not with the hypersensitive response in plant-pathogen
885 interactions. Planta. 2006;223(3):500-11.
886 57.Vlot AC, Dempsey DA, Klessig DF. Salicylic acid, a multifaceted hormone to combat disease. Annu
887 Rev Phytopathol. 2009;47:177-206.
888 58.Rowe JL, Starnes GL, Chivers PT. Complex transcriptional control links NikABCDE-dependent nickel
889 transport with hydrogenase expression in *Escherichia coli*. J Bacteriol. 2005;187(18):6317.
890 59.Biggs J. Ecology and biological control of *Agrobacterium vitis*, the grapevine crown gall pathogen.
891 Adelaide: University of Adelaide; 1994.
892 60.Konstantinidis KT, Rossello-Mora R, Amann R. Uncultivated microbes in need of their own
893 taxonomy. The ISME journal. 2017;11(11):2399-406.
894 61.Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era:
895 advancements and challenges ahead. Curr Opin Microbiol. 2007;10(5):504-9.
896 62.Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J, et al. A proposed genus boundary for the
897 prokaryotes based on genomic insights. J Bacteriol. 2014;196(12):2210.
898 63.Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson D. A genus definition for *Bacteria*
899 and *Archaea* based on a standard genome relatedness index. mBio. 2020;11(1):e02475-19.
900 64.Farrand SK, van Berkum PB, Oger P. *Agrobacterium* is a definable genus of the family
901 *Rhizobiaceae*. Int J Syst Evol Microbiol. 2003;53(5):1681-7.
902 65.Lindström K, Young JPW. International Committee on Systematics of Prokaryotes Subcommittee
903 on the taxonomy of *Agrobacterium* and *Rhizobium*: Minutes of the meeting, 7 September 2010,
904 Geneva, Switzerland. Int J Syst Evol Microbiol. 2011;61(12):3089-93.
905 66.Oren A, Garrity GM. List of new names and new combinations previously effectively, but not
906 validly, published. Int J Syst Evol Microbiol. 2016;66(11):4299-305.
907 67.Oren A, Garrity G. Notification of changes in taxonomic opinion previously published outside the
908 IJSEM. Int J Syst Evol Microbiol. 2020;70(7):4061-90.
909 68.Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
910 Bioinformatics. 2014;30(15):2114-20.
911 69.Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for
912 accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucl Acids Res.
913 2018;46(W1):W537-W44.
914 70.Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome
915 assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.
916 71.Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
917 72.Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI
918 prokaryotic genome annotation pipeline. Nucl Acids Res. 2016;44(14):6614-24.
919 73.Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional
920 characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726-31.
921 74.El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families
922 database in 2019. Nucl Acids Res. 2018;47(D1):D427-D32.
923 75.Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for
924 gene and protein annotation. Nucl Acids Res. 2016;44(Database issue):D457-D62.
925 76.Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of
926 metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucl
927 Acids Res. 2013;42(D1):D459-D71.
928 77.Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B. GET_PHYLOMARKERS, a software package to
929 select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used
930 for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. Front Microbiol. 2018;9(771).

931 78.Goris J, Konstantinidis K, Klappenbach J, Coenye T, Vandamme P, Tiedje J. DNA-DNA hybridization
932 values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol.
933 2007;57.
934 79.Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation
935 with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14(1):1-14.
936 80.Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All ANIs are not created equal:
937 implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. Int
938 J Syst Evol Microbiol. 2020;70(4):2937-48.
939 81.Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in
940 diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods.
941 2016;8(1):12-24.
942 82.Yoon S-H, Ha S-m, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate
943 average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281-6.
944 83.Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of
945 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.
946 84.Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, et al.
947 Microbial species delineation using whole genome sequences. Nucl Acids Res. 2015;43(14):6761-71.
948 85.Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the
949 prokaryotic world. Nucleic Acids Res. 2008;36(21):6688-719.
950 86.Tempé J, Petit A, Holsters M, Montagu M, Schell J. Thermosensitive step associated with transfer
951 of the Ti plasmid during conjugation: Possible relation to transformation in crown gall. Proceedings of
952 the National Academy of Sciences of the United States of America. 1977;74(7):2848-9.
953 87.Morton ER, Fuqua C. Laboratory maintenance of *Agrobacterium*. Current Protocols in
954 Microbiology. 2012;24(1):3D.1.-3D.1.6.
955 88.Schumann P, Maier T. Chapter 13 - MALDI-TOF mass spectrometry applied to classification and
956 identification of bacteria. In: Goodfellow M, Sutcliffe I, Chun J, editors. Methods Microbiol. 41:
957 Academic Press; 2014. p. 275-306.
958 89.Tóth EM, Schumann P, Borsodi AK, Kéki Z, Kovács AL, Márialigeti K. *Wohlfahrtiimonas*
959 *chitiniclastica* gen. nov., sp. nov., a new gammaproteobacterium isolated from *Wohlfahrtia magnifica*
960 (Diptera: Sarcophagidae). Int J Syst Evol Microbiol. 2008;58(Pt 4):976-81.
961 90.Bini F, Geider K, Bazzi C. Detection of *Agrobacterium vitis* by PCR using novel *virD2* gene-specific
962 primers that discriminate two subgroups. Eur J Plant Pathol. 2008;122(3):403-11.
963 91.Bini F, Kuczmog A, Putnoky P, Otten L, Bazzi C, Burr TJ, et al. Novel pathogen-specific primers for
964 the detection of *Agrobacterium vitis* and *Agrobacterium tumefaciens*. Vitis. 2008;47:181-9.
965 92.Fuller SL, Savory EA, Weisberg AJ, Buser JZ, Gordon MI, Putnam ML, et al. Isothermal amplification
966 and lateral-flow assay for detecting crown-gall-causing *Agrobacterium* spp. Phytopathol.
967 2017;107(9):1062-8.

968

969 **Tables:**

970

971 **Table 1** Characteristics of 14 strains of the *All. vitis* species complex 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[a]	CDS[a]	Accession number
K309[T]	*All. vitis sensu stricto*	Australia	1977	(9)	(48)	22	999	5.75	57.55	5188	5136	LMVL00000000.2
IPV-BO 6186	*All. vitis sensu stricto*	Italy	2006	(90)	This study	79	608	5.80	57.57	5250	5196	VOLK00000000.1
IPV-BO 7105	*All. vitis sensu stricto*	Italy	2007	(28)	This study	91	462	5.81	57.54	5214	5157	VOLJ00000000.1
KFB 239	*All. vitis sensu stricto*	Serbia	2010	(42)	This study	82	456	6.15	57.57	5545	5490	VOLI00000000.1
KFB 253	*All. vitis sensu stricto*	Serbia	2011	(42)	This study	70	401	5.81	57.56	5290	5235	VOLF00000000.1
S4[T]	*All. ampelinum*	Hungary	1981	(38)	(18)	CG[b]	CG[b]	6.32	57.47	5840	5770	CP000633.1-CP000639.1
IPV-BO 1861-5	*All. ampelinum*	Italy	1984	(91)	This study	110	971	5.67	57.63	5125	5072	VOLM00000000.1
IPV-BO 5159	*All. ampelinum*	Italy	2003	(91)	This study	160	269	6.47	57.56	5939	5883	VOLL00000000.1
KFB 243	*All. ampelinum*	Serbia	2011	(42)	This study	210	270	6.52	57.62	5963	5909	VOLH00000000.1
KFB 250	*All. ampelinum*	Serbia	2011	(42)	This study	135	356	6.47	57.60	5845	5790	VOLG00000000.1
KFB 254	*All. ampelinum*	Serbia	2011	(42)	This study	76	492	5.96	57.60	5433	5380	VOLE00000000.1

KFB 264	*All. ampelinum*	Serbia	2011	(42)	This study	75	714	5.92	57.52	5355	5299	VOLD00000000.1
V80/94	*All. ampelinum*	USA	Oregon	(92)	(92)	66	413	5.98	57.48	5467	5414	NBZE00000000.1
Av2	*Allorhizobium* sp.	Croatia	2006	(28)	This study	123	407	6.23	57.58	5713	5659	VOLN00000000.1

[a] Numbers based on Prokka annotations.

[b] CG, Complete genome.

975 **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-Butyric 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[a]	Positive
Positive tests[a]	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[a]	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

976 [a]These tests were performed for strains S4^T, KFB 243, KFB 250, KFB 254 and KFB 264 by (42), except for a test of acid
977 production in a medium containing 4-hydroxyphenylacetate conducted in this study. For strains that were not
978 included in our former study, test of production of alkali from tartrate was conducted in the present work.

979 **Figure captions:**

980 **Fig. 1** Maximum-likelihood core-genome phylogeny of 69 strains belonging to the genus
981 *Allorhizobium* and other *Rhizobiaceae* members (part collapsed). The tree was estimated with IQ-
982 TREE from the concatenated alignment of 344 top-ranked genes selected using GET_PHYLOMARKERS
983 software. The numbers on the nodes indicate the approximate Bayesian posterior probabilities
984 support values (first value) and ultra-fast bootstrap values (second value), as implemented in IQ-
985 TREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the outgroup. The scale bar
986 represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R6
987 model. "*Allorhizobium vitis*" clade is collapsed on the left tree and shown expanded on the right. The
988 matrix represents the distribution of ANIb values for the genomic sequences of the clade
989 corresponding to the *All. vitis* species complex, relative to the typical species delimitation threshold
990 of 95%. The same tree, but without collapsing clades, is presented in the Figure S1.

991 **Fig. 2** Maximum-likelihood pan-genome phylogeny of 69 strains belonging to the genus *Allorhizobium*
992 and other *Rhizobiaceae* members (part collapsed). The tree was estimated with IQ-TREE from the
993 consensus (COGtriangles and OMCL clusters) gene presence/absence matrix containing 33,396
994 clusters obtained using GET_HOMOLOGUES software. The numbers on the nodes indicate the
995 approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap
996 values (second value), as implemented in IQ-TREE. The tree was rooted using the *Mesorhizobium*
997 spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per
998 site under the best-fitting GTR2+FO+R5 model. The same tree, but without collapsing clades, is
999 presented in the Figure S3.

1000 **Fig. 3** Pan-genome analyses of 14 *All. vitis* species complex strains. **a)** Bar plot showing the
1001 frequencies of orthologous clusters as predicted by the COGtriangles and OMCL algorithms. **b)** Pie
1002 chart showing the relative sizes (cluster numbers) contained in the core, soft-core, shell, and cloud
1003 genome compartments.

1004 **Supplementary Information**

1005

1006 **Additional file 1: Fig. S1.** Maximum-likelihood core-genome phylogeny of 69 strains belonging to the
1007 genus *Allorhizobium* and other *Rhizobiaceae* members (uncollapsed). The tree was estimated with
1008 IQ-TREE from the concatenated alignment of 344 top-ranked genes selected using
1009 GET_PHYLOMARKERS software. The numbers on the nodes indicate the approximate Bayesian
1010 posterior probabilities support values (first value) and ultra-fast bootstrap values (second value), as
1011 implemented in IQ-TREE. The tree was rooted using the *Mesorhizobium* spp. sequences as the
1012 outgroup. The scale bar represents the number of expected substitutions per site under the best-
1013 fitting GTR+F+ASC+R6 model. The same tree, but with collapsed clades, is presented in Figure 1.

1014

1015 **Additional file 2: Fig. S2.** Maximum-likelihood core-genome phylogeny of 103 strains belonging to
1016 the genus *Allorhizobium* (including 34 additional strains of *All. vitis* species complex strains whose
1017 sequences are available in GenBank but not associated to a published study) and other *Rhizobiaceae*
1018 members. The tree was estimated with IQ-TREE from the concatenated alignment of 302 top-ranked
1019 genes selected using GET_PHYLOMARKERS software. The numbers on the nodes indicate the
1020 approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap
1021 values (second value), as implemented in IQ-TREE. The tree was rooted using the *Mesorhizobium*
1022 spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per
1023 site under the best-fitting GTR+F+ASC+R7 model. The matrix in the top-right corner represents the
1024 distribution of ANIb values for genomic sequences of the clade corresponding to the *All. vitis* species
1025 complex, relative to the typical species delimitation threshold of 95%.

1026

1027 **Additional file 3: Fig. S3.** Maximum-likelihood pan-genome phylogeny of 69 strains belonging to the
1028 genus *Allorhizobium* and other *Rhizobiaceae* members (uncollapsed). The tree was estimated with
1029 IQ-TREE from the consensus (COGtriangles and OMCL clusters) pan-genome matrix containing 33,396
1030 clusters obtained using GET_HOMOLOGUES software. The numbers on the nodes indicate the

1031 approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap
1032 values (second value), as implemented in IQ-TREE. The tree was rooted using the *Mesorhizobium*
1033 spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per
1034 site under the best-fitting GTR2+FO+R5 model. The same tree, but with collapsed clades, is presented
1035 in Figure 2.
1036
1037 **Additional file 4: Fig. S4.** Heatmap representation of the average nucleotide identity (ANIb) for TAR
1038 regions of *All. vitis* species complex strains. PyANI program Version 0.2.9
1039 (https://github.com/widdowquinn/pyani) was used to calculate ANIb values and generate the
1040 clustered heatmap.
1041
1042 **Additional file 5: Fig. S5.** Score-oriented dendrogram showing the similarity of the MALDI-TOF mass
1043 spectra of 14 *All. vitis* species complex strains studied. The dendrogram was created using the MALDI
1044 Biotyper Compass Explorer software (Bruker, Version 4.1.90).
1045
1046 **Additional file 6: Table S1.** List of additional strains and GenBank/EMBL/DDBJ accession numbers for
1047 their nucleotide sequences used in this study. **a)** List of 55 reference *Rhizobiaceae* strains and
1048 GenBank/EMBL/DDBJ accession numbers for their nucleotide sequences used in this study; **b)** List of
1049 additional 34 *All. vitis* species complex strains and GenBank/EMBL/DDBJ accession numbers for their
1050 nucleotide sequences used in this study. Although available in the public nucleotide sequence
1051 databases, these genome sequences have not yet been presented in peer-reviewed study by
1052 sequence depositors.
1053
1054 **Additional file 7: Table S2.** Pairwise OGRI comparisons amongst 14 *Allorhizobium vitis* species
1055 complex strain genomes towards species delimitation. ANI values were computed with various
1056 implementation of the ANI metric: **a)** ANIb, **b)** ANIm, **c)** orthoANI and **d)** fastANI; **e)** dDDH values
1057 were computed with the GGDC.

1058

1059 **Additional file 8: Table S3.** Pairwise ANIb comparisons amongst extended set of 48 *Allorhizobium*
1060 *vitis* species complex strain genomes towards species delimitation.

1061

1062 **Additional file 9: Table S4.** Clusters of contiguous clade-specific genes. Clusters were identified
1063 amongst sets of genes deemed specific of the focal clade based on detection by either Pantagruel or
1064 GET_HOMOLOGUES pipelines. **a)** Clusters of genes specific to *Allorhizobium vitis* species complex
1065 (present in all *All. vitis sensu stricto*, *All. ampelinum* and *Allorhizobium* sp. Av2, and in no other
1066 *Allorhizobium* spp.); **b)** Clusters of genes specific to *All. vitis sensu stricto* (present in all five tested
1067 strains and in none of *All. ampelinum*); **c)** Clusters of genes specific to *All. ampelinum* (present in all
1068 eight tested strains and in none of *All. vitis sensu stricto*).

1069

1070 **Additional file 10: Table S5.** Pairwise ANIb values between tartrate utilization (TAR) regions of
1071 *Allorhizobium vitis* species complex strains.

1072

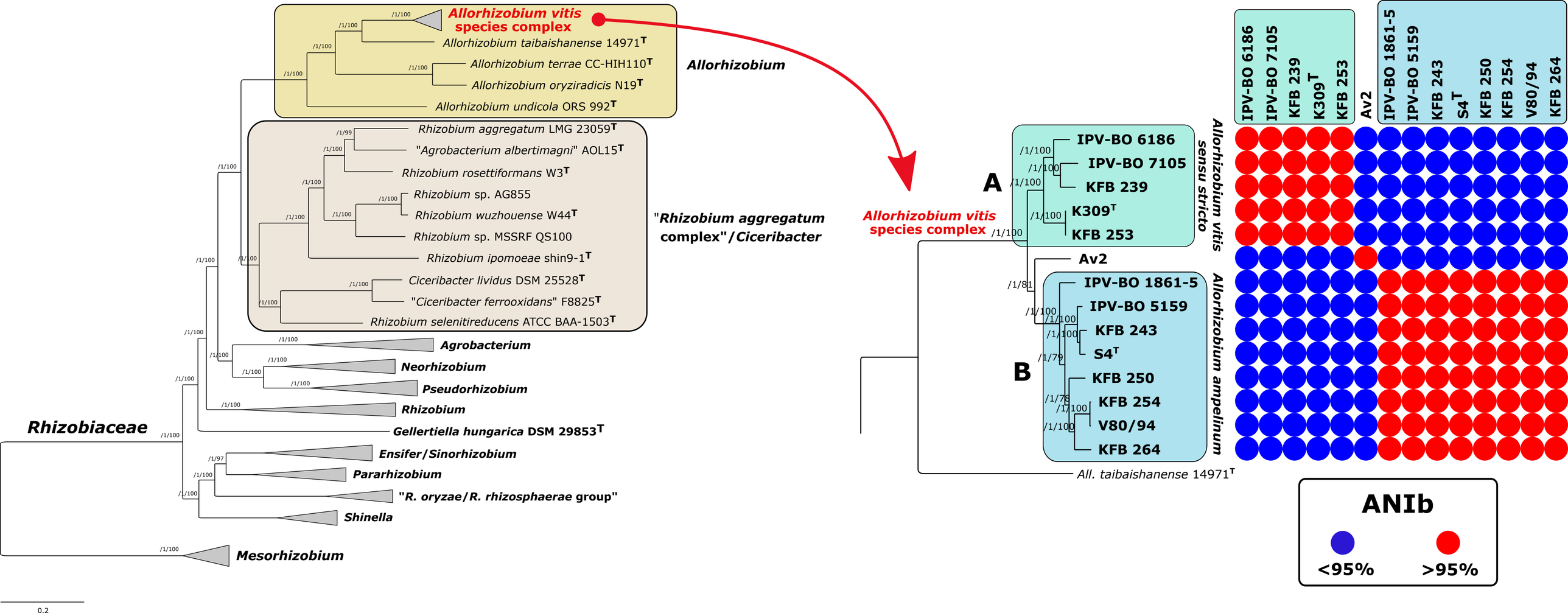
1073 **Additional file 11: Table S6.** Tartrate utilization (TAR) region genotype and tartrate metabolism
1074 phenotype (production of alkali from L-tartaric acid) of *All. vitis* species complex strains.

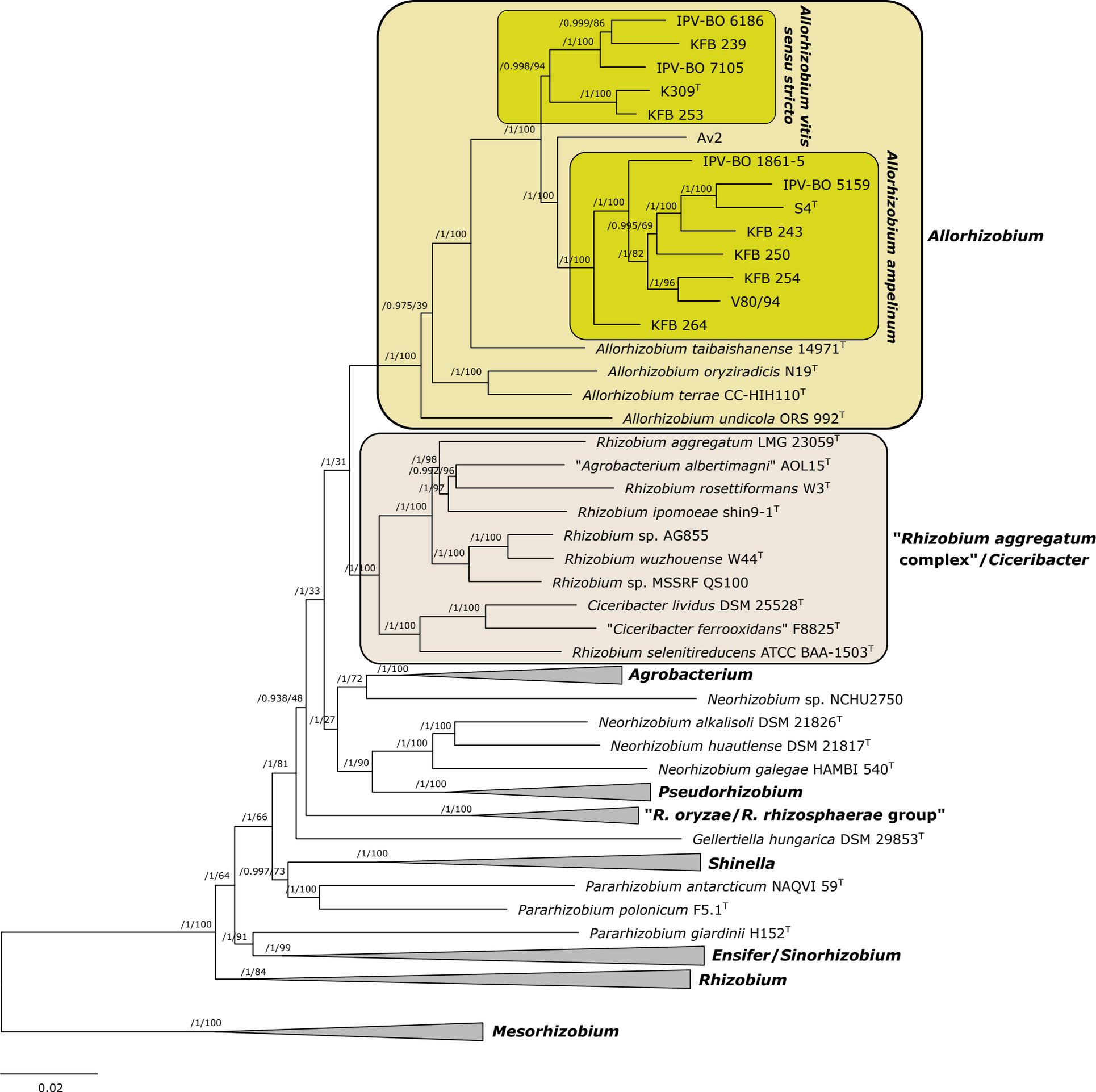
1075

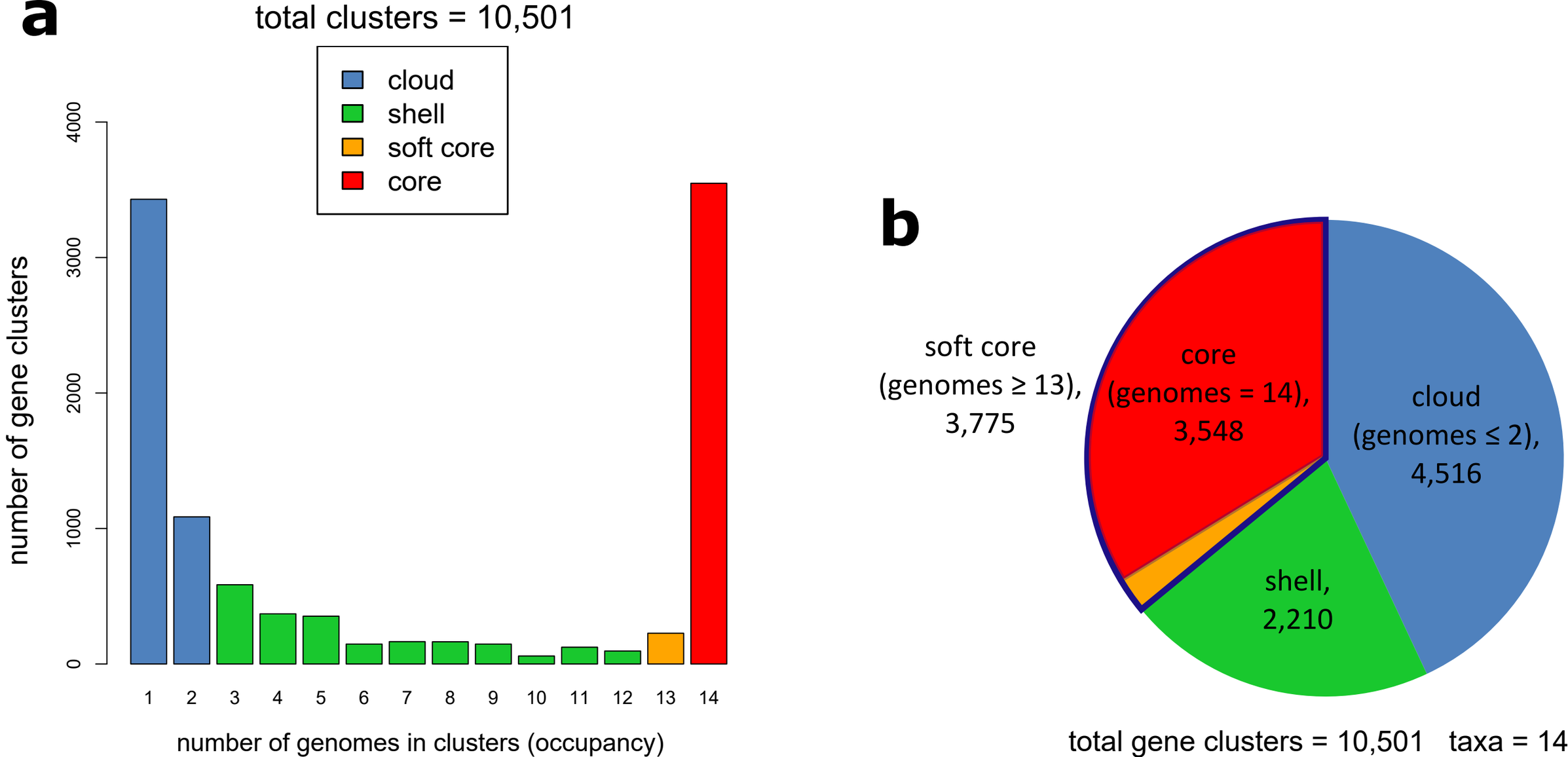
1076 **Additional file 12: Table S7.** Pairwise OGRI comparisons amongst 69 *Rhizobiaceae* strain genomes
1077 towards genus delimitation. **a)** Average amino acid identity (AAI); **b)** percentage of conserved
1078 proteins (POCP); **c)** genome-wide average nucleotide identity (gANI) and alignment fraction (AF), with
1079 AF values indicated in parentheses.

1080

1081 **Additional file 13: Supplementary Methods.** This file contains additional detailed methods
1082 description.


Allorhizobium vitis species complex
Allorhizobium taibaishanense 14971T
Allorhizobium terrae CC-HIH110T
Allorhizobium oryziradicis N19T
Allorhizobium undicola ORS 992T
Allorhizobium
Rhizobium aggregatum LMG 23059T
"Agrobacterium albertimagni" AOL15T
Rhizobium rosettiformans W3T
Rhizobium sp. AG855
Rhizobium wuzhouense W44T
Rhizobium sp. MSSRF QS100
Rhizobium ipomoeae shin9-1T
Ciceribacter lividus DSM 25528T
"Ciceribacter ferrooxidans" F8825T
Rhizobium selenitireducens ATCC BAA-1503T
"Rhizobium aggregatum complex"/Ciceribacter
Agrobacterium
Neorhizobium
Pseudorhizobium
Rhizobium
Gellertiella hungarica DSM 29853T
Ensifer/Sinorhizobium
Pararhizobium
"R. oryzae/R. rhizosphaerae group"
Shinella
Rhizobiaceae
Mesorhizobium
0.2
Allorhizobium vitis species complex
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
"Agrobacterium albertimagni" AOL15T
Rhizobium rosettiformans W3T
Rhizobium ipomoeae shin9-1T
Rhizobium sp. AG855
Rhizobium wuzhouense W44T
Rhizobium sp. MSSRF QS100
Ciceribacter lividus DSM 25528T
"Ciceribacter ferrooxidans" F8825T
Rhizobium selenitireducens ATCC BAA-1503T
"Rhizobium aggregatum complex"/Ciceribacter
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/Sinorhizobium
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