

# Improved genome of *Agrobacterium radiobacter* type strain provides new taxonomic insight into *Agrobacterium* genomospecies 4

Han Ming Gan<sup>1,2,3</sup>, Melvin V.L. Lee<sup>3</sup> and Michael A. Savka<sup>4</sup>

<sup>1</sup> Deakin Genomics Centre, Deakin University, Geelong, VIC, Australia

<sup>2</sup> Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, Australia

<sup>3</sup> School of Science, Monash University Malaysia, Petaling Jaya, Selangor, Malaysia

<sup>4</sup> College of Science, The Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY, USA

## ABSTRACT

The reported *Agrobacterium radiobacter* DSM 30174<sup>T</sup> genome is highly fragmented, hindering robust comparative genomics and genome-based taxonomic analysis. We re-sequenced the *Agrobacterium radiobacter* type strain, generating a dramatically improved genome with high contiguity. In addition, we sequenced the genome of *Agrobacterium tumefaciens* B6<sup>T</sup>, enabling for the first time, a proper comparative genomics of these contentious *Agrobacterium* species. We provide concrete evidence that the previously reported *Agrobacterium radiobacter* type strain genome (Accession Number: [ASXY01](#)) is contaminated which explains its abnormally large genome size and fragmented assembly. We propose that *Agrobacterium tumefaciens* be reclassified as *Agrobacterium radiobacter* subsp. *tumefaciens* and that *Agrobacterium radiobacter* retains its species status with the proposed name of *Agrobacterium radiobacter* subsp. *radiobacter*. This proposal is based, first on the high pairwise genome-scale average nucleotide identity supporting the amalgamation of both *Agrobacterium radiobacter* and *Agrobacterium tumefaciens* into a single species. Second, maximum likelihood tree construction based on the concatenated alignment of shared genes (core genes) among related strains indicates that *Agrobacterium radiobacter* NCPPB3001 is sufficiently divergent from *Agrobacterium tumefaciens* to propose two independent sub-clades. Third, *Agrobacterium tumefaciens* demonstrates the genomic potential to synthesize the L configuration of fucose in its lipid polysaccharide, fostering its ability to colonize plant cells more effectively than *Agrobacterium radiobacter*.

**Subjects** Agricultural Science, Genomics, Microbiology, Plant Science, Taxonomy

**Keywords** Type strain, Average nucleotide identity, Phylogenomics, *Agrobacterium radiobacter*, *Agrobacterium tumefaciens*, Lipopolysaccharide, *Agrobacterium*, Ti plasmid

## INTRODUCTION

The taxonomy and phylogeny of the genus *Agrobacterium* has proven to be complex and controversial. Bacteria of the genus *Agrobacterium* have been grouped into six species

Submitted 23 October 2018  
Accepted 20 December 2018  
Published 8 February 2019

Corresponding author  
Michael A. Savka, [massbi@rit.edu](mailto:massbi@rit.edu)

Academic editor  
Mikhail Gelfand

Additional Information and  
Declarations can be found on  
page 15

DOI [10.7717/peerj.6366](https://doi.org/10.7717/peerj.6366)

© Copyright  
2019 Gan et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

based on the disease phenotype associated, in part, with the resident disease-inducing plasmid. Among those six species are *Agrobacterium tumefaciens* causing crown gall on dicotyledonous plants, stone fruit and nut trees and *Agrobacterium radiobacter* that is not known to cause plant diseases of any kind (Bouzar & Jones, 2001; Conn, 1942; Kerr & Panagopoulos, 1977; Panagopoulos, Psallidas & Alivizatos, 1978; Riker et al., 1930; Starr & Weiss, 1943; Süle, 1978). An alternative classification approach grouped *Agrobacterium* organisms into three biovars based on physiological and biochemical properties without consideration of disease phenotype (Keane, Kerr & New, 1970; Kerr & Panagopoulos, 1977; Panagopoulos, Psallidas & Alivizatos, 1978). The species and biovar classification schemes do not coincide well, in a large part, because of the disease-inducing plasmids, tumor-inducing (pTi) and hairy root-inducing (pRi), are readily transmissible plasmids (Young et al., 2001).

Many widely used approaches for bacterial species definition include composition of peptidoglycan, base composition of DNA, fatty acid and 16S rDNA sequence (Stackebrandt et al., 2002) in addition to newer methods based on the whole-genome analysis (Coutinho et al., 2016; Jain et al., 2018), horizontal gene transfer analysis (Bobay & Ochman, 2017) or the core genome analysis (Moldovan & Gelfand, 2018) which is used in the present study. The genus *Agrobacterium* is a prime example with many proposals and oppositions regarding the amalgamation of *Agrobacterium* and *Rhizobium* over the last three or four decades (Farrand, Van Berkum & Oger, 2003; Gaunt et al., 2001; Young et al., 2001, 2003). However, more recent studies appear to favor the preservation of the genus *Agrobacterium* backed by strong genetic and genomic evidence (Gan & Savka, 2018; Ramírez-Bahena et al., 2014). Within the genus *Agrobacterium*, the taxonomic status of *Agrobacterium radiobacter* and *Agrobacterium tumefaciens* remains contentious (Sawada et al., 1993; Young, 2008; Young, Pennycook & Watson, 2006). *Agrobacterium radiobacter* (originally proposed as *Bacillus radiobacter*) is a non-pathogenic soil bacterium associated with nitrogen utilization isolated more than a century ago in 1902 (Beijerinck & Van Delden, 1902; Conn, 1942). On the other hand, *Agrobacterium tumefaciens* (previously *Bacterium tumefaciens*) is a plant pathogen capable of inducing tumorigenesis (Smith & Townsend, 1907). However, the descriptive assignment for *Agrobacterium tumefaciens* was later found to be contributed by a set of genes located on the large Ti plasmid that can be lost (Gordon & Christie, 2014). In other words, the curing of Ti plasmid in *Agrobacterium tumefaciens* will change its identity to the non-pathogenic species, *Agrobacterium radiobacter*. Furthermore, comparative molecular analysis based on single-copy housekeeping genes also supports the close relatedness of *Agrobacterium radiobacter* and *Agrobacterium tumefaciens*, blurring the taxonomic boundaries between these species (Mousavi et al., 2015; Shams et al., 2013). As taxa are reclassified into different populations that do not conform to the characteristics of the original description, the given names lose their significant and descriptive importance. Consistent with the Judicial Commission according to the Rules of the International Code of Nomenclature of Bacteria, Tindall (2014) concluded that the combination of *Agrobacterium radiobacter* has priority over the combination *Agrobacterium tumefaciens* when the two are treated as members of the same species since *Agrobacterium*

*radiobacter* was the first proposed and described in 1902 whereas *Agrobacterium tumefaciens* was first proposed and described in 1907) (Tindall, 2014). However, given that *Agrobacterium tumefaciens* has been more widely studied than *Agrobacterium radiobacter* due to its strong relevance to agriculture (Bourras, Rouxel & Meyer, 2015), it remains unclear but interesting to see if the broader scientific community will obey this rule by adopting the recommended species name change in future studies.

To our knowledge, a detailed comparative genomics analysis of *Agrobacterium radiobacter* and *Agrobacterium tumefaciens* type strains has not been reported despite their genome availability (Zhang et al., 2014). The high genomic relatedness of both type strains was briefly mentioned by Kim & Gan (2017) through whole genome alignment and pairwise nucleotide identity calculation from homologous regions. However, evidence is now mounting that the *Agrobacterium radiobacter* DSM 30147<sup>T</sup> reported by Zhang et al. (2014) is contaminated, warranting immediate investigation (Jeong, Pan & Park, 2016). The assembled genome is nearly 7 megabases, the largest among *Agrobacterium* currently sequenced at that time with up to 6,853 predicted protein-coding genes contained in over 600 contigs. At sequencing depth of nearly 200×, its genome assembly is unusually fragmented even for a challenging microbial genome (Utturkar et al., 2017). Furthermore, the phylogenomic placement of *Agrobacterium radiobacter* DSM 30147<sup>T</sup> based on this genome assembly has been questionable as evidenced by its basal position and substantially longer branch length relative to other members of the species (Gan & Savka, 2018). The overly fragmented nature of this assembly also precludes fruitful comparative genomics focusing on gene synteny analysis. More importantly, analysis done on a contaminated assembly but with the assumption that it is not, will likely lead to incorrect biological interpretations (Allnutt et al., 2018).

In this study, we sequenced the whole genome of *Agrobacterium radiobacter* using a type strain that was sourced from the National Collection of Plant Pathogenic Bacteria (NCPBB). We produced a contiguous genome assembly exhibiting genomic statistics that are more similar to other assembled *Agrobacterium* genomes. We show here, through comparative genomics and phylogenetics, that the previously assembled *Agrobacterium radiobacter* DSM 30147<sup>T</sup> genome contains substantial genomic representation from another *Agrobacterium* sp. isolated and sequenced by the same lab, consistent with our initial suspicion of strain contamination. Using the newly assembled genome for subsequent comparative analysis, we provide genomic evidence that *Agrobacterium radiobacter* DSM 30147<sup>T</sup> and *Agrobacterium tumefaciens* B6<sup>T</sup> are the same species. However, strain DSM 30147<sup>T</sup> should not be considered as a merely non-tumorigenic strain of *Agrobacterium tumefaciens* as substantial genomic variation exists between these two type strains notably in the nucleotide sugar metabolism pathway that may contribute to their ecological niche differentiation.

## MATERIALS AND METHODS

### DNA extraction and whole genome sequencing

Approximately 10 bacterial colonies were scrapped using a sterile P200 pipette tip from a 3-day-old nutrient agar culture and resuspended in lysis buffer with proteinase K

([Sokolov, 2000](#)) followed by incubation at 56 °C for 3 h. DNA purification was performed as previously described. The extracted DNA was normalized to 0.2 ng/μL and prepared using the Nextera XT library preparation kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The library was sequenced on an Illumina MiSeq desktop sequencer located at the Monash University Malaysia Genomics Facility (2 × 250 bp run configuration) that routinely sequences mostly decapod crustacean mitogenomes ([Gan, Tan & Austin, 2016a](#); [Gan et al., 2016b](#); [Tan et al., 2015](#)) and occasionally microbial genomes ([Gan et al., 2014, 2015](#); [Wong et al., 2014](#)) without prior history of processing any member from the *Agrobacterium* genomospecies 4.

### De novo assembly and genome completeness assessment

Raw paired-end reads were adapter-trimmed using Trimmomatic v0.36 ([Bolger, Lohse & Usadel, 2014](#)) followed by error-correction and de novo assembly using Spades Assembler v3.9 ([Bankevich et al., 2012](#)) (See [Data S1](#) for specific trimming and assembly settings). Genome completeness was assessed with BUSCOv3 (Rhizobiales database) ([Waterhouse et al., 2017](#)).

### Protein clustering

Gene prediction used Prodigal v2.6 ([Hyatt et al., 2010](#)). Clustering of the predicted coding sequence was performed with CD-HIT-EST using the settings “-C 0.95, -T 0.8” ([Li & Godzik, 2006](#)). Identification of unique and shared clusters were done using basic unix commands, for example, csplit, grep, sort and uniq. The specific commands used and files generated during clustering can be found in the Zenodo database (<https://doi.org/10.5281/zenodo.1489356>).

### Phylogenetic analysis

Reconstruction of the *Agrobacterium* phylogeny used PhyloPhlAN ([Segata et al., 2013](#)). PhyloPhlAN is a bioinformatic pipeline that identifies conserved proteins (400 markers) from microbial genomes and uses them to construct a high-resolution phylogeny using maximum likelihood inference approach ([Price, Dehal & Arkin, 2010](#)). For single gene tree construction, protein sequences were aligned with mafft v7.3 ([Katoh & Standley, 2013](#)) using the the most accurate setting (-localpair -maxiterate 1000) followed by phylogenetic tree construction via IqTree v1.65 with optimized model ([Kalyaanamoorthy et al., 2017](#); [Nguyen et al., 2014](#)). Visualization and annotation of phylogenetic trees was performed with Figtree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

### Pan-genome construction and phylogenomics

Whole genome sequences were reannotated with Prokka v1.1 using the default setting ([Seemann, 2014](#)). The Prokka-generated gff files were used as the input for Roary v3.12.0 to calculate the pan-genome ([Page et al., 2015](#)). Maximum likelihood tree construction of the core-genome alignment and tree visualization used FastTree2 v2.1.10 (-nt -gtr) ([Price, Dehal & Arkin, 2010](#)) and FigTree v 1.4.3, respectively. Input and output files associated with the Roary analysis have been deposited in the Zenodo database (<https://doi.org/10.5281/zenodo.1489356>).

## Detection and visualization of Ti plasmid

Genome sequences of each member of the genomospecies 4 except for the problematic DSM 37014<sup>T</sup> strain were used as the query for blastN search (*e*-value  $1e-100$ ) against the octopine-type Ti plasmid (Altschul *et al.*, 1990). The result of the similarity search was subsequently visualized in Blast Ring Image Generator v0.95 (Alikhan *et al.*, 2011).

## Genome annotation and KEGG pathway reconstruction

Whole genome sequences of *Agrobacterium tumefaciens* B6<sup>T</sup> and *Agrobacterium radiobacter* NCPPB 3001<sup>T</sup> were submitted to the online server GhostKoala (Kanehisa, Sato & Morishima, 2016b) for annotation and the annotated genomes were subsequently used to reconstruct KEGG pathways (Kanehisa *et al.*, 2016a) in the same webserver. Identification of proteins with TIGRFAM signatures of interest (Haft, Selengut & White, 2003) used HMMsearch v3.1b2 with the option “-cut\_tc” activated to filter for only protein hits passing the TIGRFAM trusted cutoff values (Johnson, Eddy & Portugaly, 2010).

## RESULTS

### An improved *Agrobacterium radiobacter* type strain genome

Raw sequencing data and whole genome assembly for strains B6 and NCPPB3001 reported in this study are linked to the NCBI Bioproject IDs PRJNA300485 and PRJNA300611, respectively. The newly assembled genome of *Agrobacterium radiobacter* type strain that was sourced from the NCPPB is approximately 30% smaller than the first reported *Agrobacterium radiobacter* DSM 30147<sup>T</sup> genome with 96% less contigs (22 vs 612), 20-fold longer N50 (480 vs 23 kb) and assembled length that is much more similar to other *Agrobacterium* spp. (Table 1). In addition, it is near-complete with 685 out of 686 BUSCO Rhizobiales single-copy genes detected as either partial or complete with minimal evidence of contamination as indicated by the near absence of duplicated single-copy gene (<0.1%). On the contrary, the current DSM 30147 genome is missing 25.1% of the single copy gene with up to 34.8% duplication rate. At the time of this manuscript writing, another genome of *Agrobacterium radiobacter* type strain that was sourced from another culture collection centre, for example, the Belgian Coordinated Collections of Microorganisms has been deposited in the NCBI wgs database (*Agrobacterium radiobacter* LMG140<sup>T</sup>; Table 1) with assembly statistics that are highly similar to the type strain genome reported in this study.

### The inflated genome size of *Agrobacterium radiobacter* DSM 30147(T) is due to technical errors

Instead of sharing a recent common ancestor as would be expected for a recently duplicated gene, the duplicated single copy genes coding for seryl-tRNA synthetase in *Agrobacterium radiobacter* DSM 30147<sup>T</sup> were placed in two distinct clusters with one affiliated to genomospecies 4 and the other affiliated to genomospecies 7 (Fig. 1A). Such an unexpected clustering pattern raises the suspicion of genome assembly from two or more non-clonal bacterial strains. In addition, by performing comparison at the

**Table 1** Genome statistics of publicly available *Agrobacterium* genomospecies 4 whole genome sequences.

Assembly accession	Strain	Isolation source	Country	Size	GC%	# Contig
<a href="#">GCF_900045375</a>	B6	Apple Gall (Iowa)	USA	5.8	59.07	4
<a href="#">GCF_001541315*</a>	B6	Apple Gall (Iowa)	USA	5.6	59.32	52
<a href="#">GCF_001692245</a>	B140/95	Peach/Almond Rootstock	USA	5.7	59.23	45
<a href="#">GCF_002179795</a>	LMG 215	<i>Humulus lupulus</i> gall (USA)	USA	5.4	59.48	33
<a href="#">GCF_000233975</a>	CCNWGS0286	<i>R. pseudoacacia</i> nodules	China	5.2	59.53	49
<a href="#">GCF_900011755</a>	Kerr 14 = LMG 15 = CFBP 5761	Soil around <i>Prunus dulcis</i>	Australia	5.9	59.04	5
<a href="#">GCF_002591665</a>	186	English Walnut gall	California	5.7	59.42	22
<a href="#">GCF_002008215</a>	LMG 140 = NCPPB 3001 = CFBP 5522 = DSM 30147	Saprobic soil	Germany	5.5	59.34	22
<a href="#">GCF_000421945</a>	LMG 140 = NCPPB 3001 = CFBP 5522 = DSM 30147	Saprobic soil	Germany	7.17	59.86	612
<a href="#">GCF_001541305*</a>	LMG 140 = NCPPB 3001 = CFBP 5522 = DSM 30147	Saprobic soil	Germany	5.5	59.36	22
<a href="#">GCF_900012605</a>	CFBP 5621	<i>Lotus corniculata</i> , root tissue commensal	France	5.4	59.32	3
<a href="#">GCF_003031125</a>	LAD9 (CGMCC No. 2962)	Landfill leachate treatment system	China	5.9	59.13	49
<a href="#">GCF_000384555</a>	224MFTsu31	Rhizosphere of <i>L. luteus</i> in Hungary, formerly <i>R. lupini</i> H13-3	USA	4.8	59.73	21
<a href="#">GCF_900188475</a>	719_389	Rhizosphere and endosphere of <i>Arabidopsis thaliana</i> .	USA	4.9	59.73	18
<a href="#">GCF_000384555</a>	UNC420CL41Cvi	Plant associated	USA	5	59.69	18

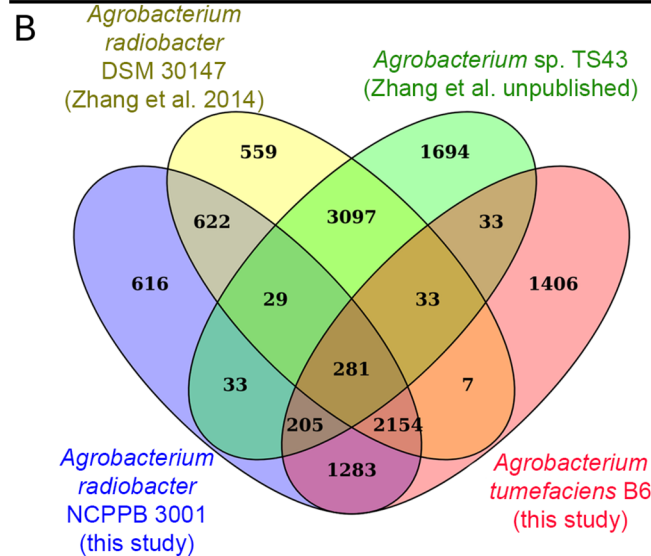
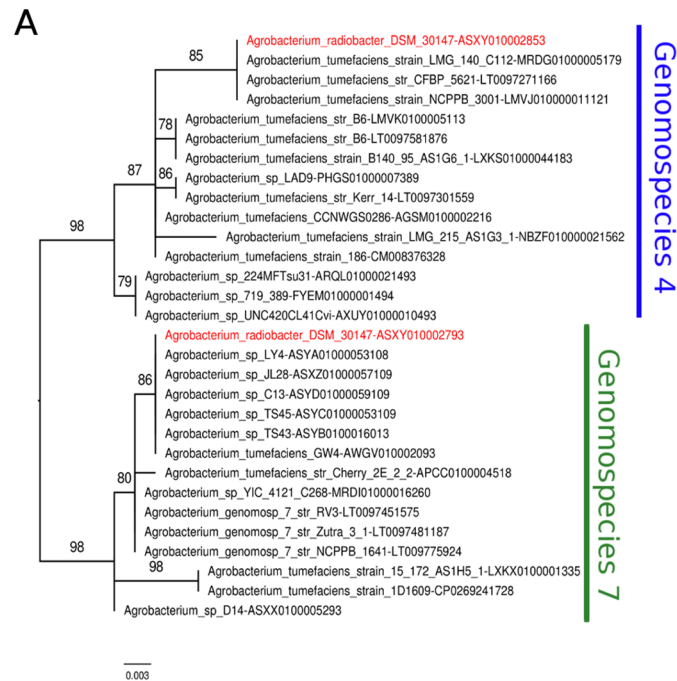
**Note:**

\* Reported in this study.

genome-scale based on whole proteome clustering of *Agrobacterium radiobacter* DSM 30147<sup>T</sup>/NCPPB 3001<sup>T</sup> (Previous study, [GCF\\_000421945](#); This study, [GCF\\_001541305](#)), *A. sp.* TS43 (unpublished, [GCF\\_001526605](#)) and *Agrobacterium tumefaciens* B6 ([GCF\\_001541315](#)), we observed a high number of proteins that were exclusively shared between Zhang et al. *Agrobacterium radiobacter* DSM 30147 and *A. sp.* TS43 belonging to genomospecies 7 ([Fig. 1B](#)). Coincidentally, despite not sharing the same Bioproject ID, the whole genomes of strains DSM 30147<sup>T</sup> and TS43 were sequenced by the Zhang et al., and submitted to NCBI on the same date, 30 May 2013, hinting strain contamination during sample processing in the lab.

### Genome-scale average nucleotide identity calculation supports the amalgamation of *Agrobacterium radiobacter* and *Agrobacterium tumefaciens* into a single genomospecies

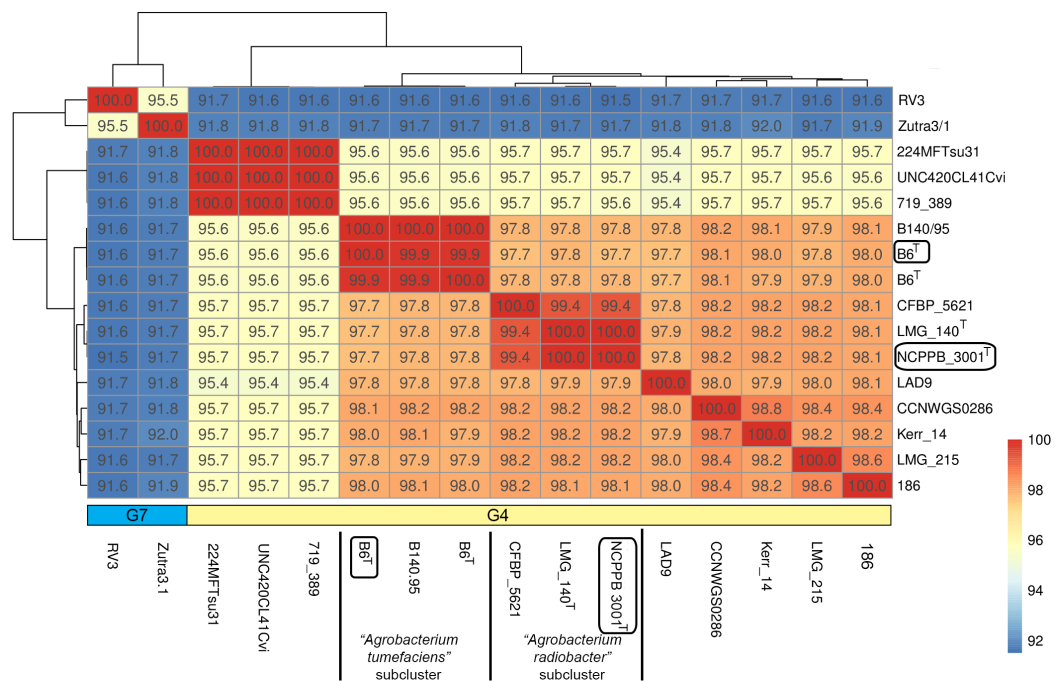
Single gene tree shows that *Agrobacterium radiobacter* NCPPB 3001<sup>T</sup> and *Agrobacterium tumefaciens* B6<sup>T</sup> belong to the genomospecies 4 clade ([Fig. 1A](#)), corroborating with the PhyloPhlAN phylogenomic tree that was constructed based on the alignment of



**Figure 1** Phylogenetic and genomic evidence indicating contamination in the published *A. radiobacter* DSM 30147T genome. (A) Maximum likelihood phylogenetic tree of seryl-tRNA synthetases from *Agrobacterium* genomespecies 4 and 7. Codes after the tildes are contigs containing the corresponding homologs. Node labels indicate ultra-fast bootstrap support value and branch length indicates number of substitutions per site. Duplicated homologs in the problematic *A. radiobacter* DSM 30147 genome were colored red. (B) Venn diagram of the core proteome of selected *Agrobacterium* strains from genomespecies 4. Numbers in the overlapping regions indicate the number of coding sequences (CDS) that shared by two or more groups at 95% nucleotide identity cutoff.

Full-size DOI: 10.7717/peerj.6366/fig-1

400 universal single-copy proteins (Fig. S1). The pairwise average nucleotide identity (ANI) among strains within this clade is consistently more than 95% further supporting their affiliation to the same genomespecies (Fig. 2) (Coutinho et al., 2016; Jain et al., 2018).



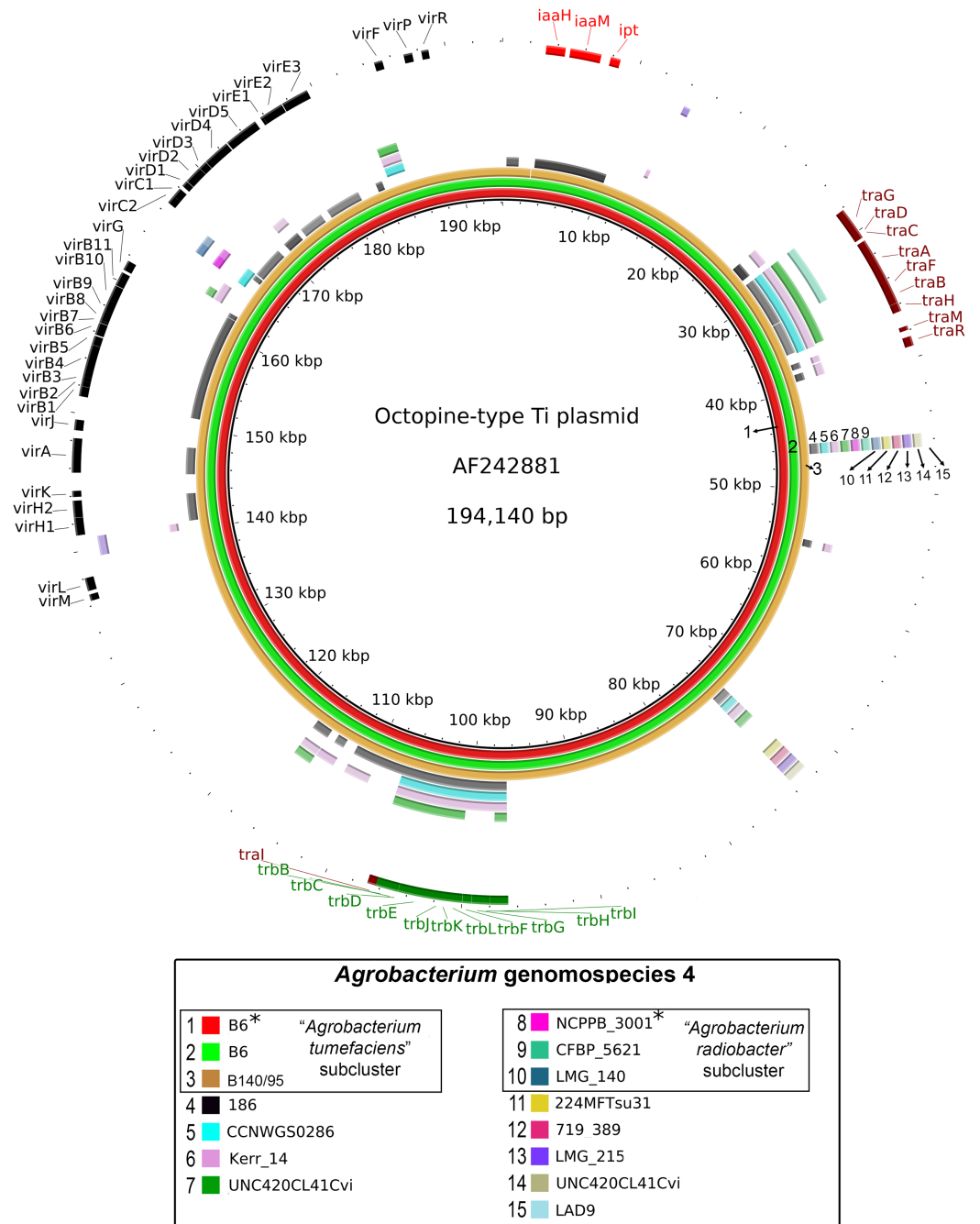
**Figure 2** A heatmap showing the hierarchical clustering of *Agrobacterium* strains based on genomic distance. Values in boxes indicate pairwise average nucleotide identity. Horizontal colored bar below the heatmap indicate the genomespecies assigned to each genome (G7, genomespecies 7; G4, genomespecies 4). Boxed labels indicate genomes sequenced in this study. [Full-size DOI: 10.7717/peerj.6366/fig-2](https://doi.org/10.7717/peerj.6366/fig-2)

As expected, pairwise ANI of less than 92% was observed when they were compared with strains from genomespecies 7 (strains RV3 and Zutra 3/1). A 100% pairwise ANI was observed between *Agrobacterium radiobacter* type strains that were sourced from NCPPB and LMG. In addition, non-type strains B140/95 and CFBP5621 also exhibit a strikingly high pairwise ANI (>99%) to the type strains of *Agrobacterium tumefaciens* and *Agrobacterium radiobacter*, respectively, leading to the formation of sub-clusters within genomespecies 4 (Fig. 2).

### Is *Agrobacterium radiobacter* a non-tumorigenic strain of *Agrobacterium tumefaciens*?

A majority of the currently sequenced strains from genomespecies 4 are non-tumorigenic as evidenced by the near complete lack of genomic region with significant nucleotide similarity to the octopine-type Ti reference plasmid (Fig. 3). Of the 14 genomes analyzed, only strains B6<sup>T</sup> and B140/95 exhibit a complete coverage of the Ti plasmid with near 100% sequence identity while strain 186 shows hits mainly to the essential gene clusters of a Ti plasmid such as the *vir* gene cluster (black rings and gene labels in Fig. 3) at a substantially lower sequence identity (50% <  $x$  < 90%) (Fig. 3), suggesting that it may be harboring a dissimilar variant of Ti plasmid, for example, different opine type. In addition, although lacking hits to the virulence gene of the Ti plasmid, the *tra* and *trb* clusters involved in plasmid conjugal transfer are present in strains Kerr 14, CCNWGS0286 and UNC420CL41Cvi. Despite belonging to the same genomespecies,



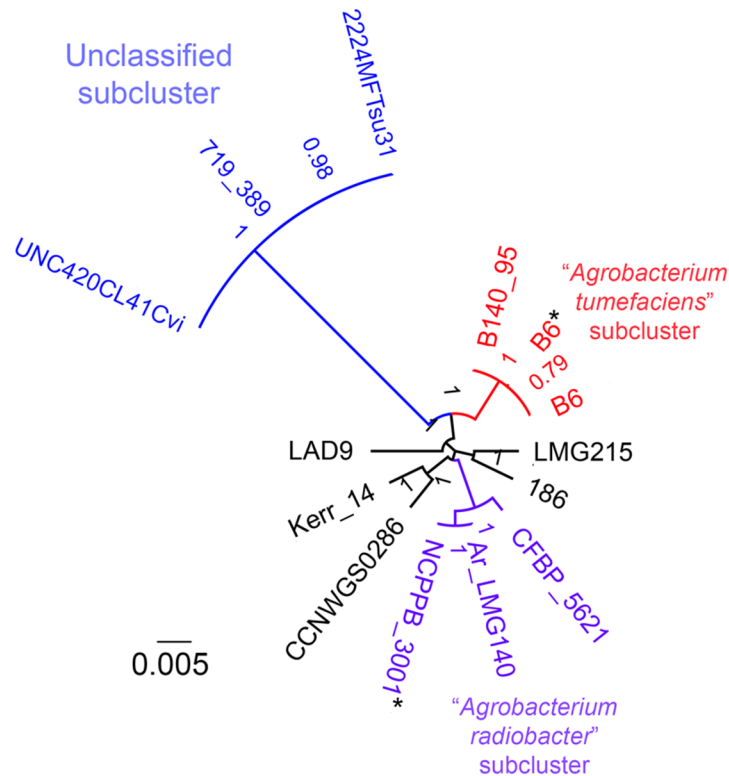


**Figure 3** Prevalence and sequence conservation of the octopine-type Ti plasmid among *Agrobacterium* genomespecies 4. Each genome (labelled 1–15) is represented by a colored ring shaded based on nucleotide percentage similarity to the reference Ti plasmid (min. 50%; max. 100%). The outermost ring highlights the gene regions involved in tumorigenesis (*vir*, *iaa* and *ipt*) and plasmid conjugation (*trb* and *tra*). Asterisks indicate genomes sequenced in this study.

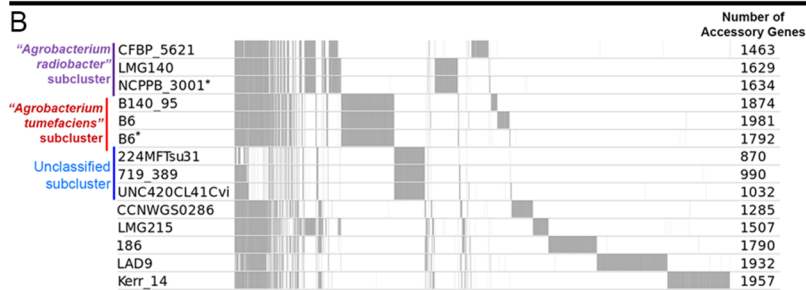
Full-size DOI: 10.7717/peerj.6366/fig-3

core genome alignment and phylogenomic analysis indicates that *Agrobacterium radiobacter* NCPBP3001<sup>T</sup> is sufficiently divergent from *Agrobacterium tumefaciens* B6<sup>T</sup> leading to their separation into two distinct sub-clusters (Fig. 4A). This is also resonated by

A

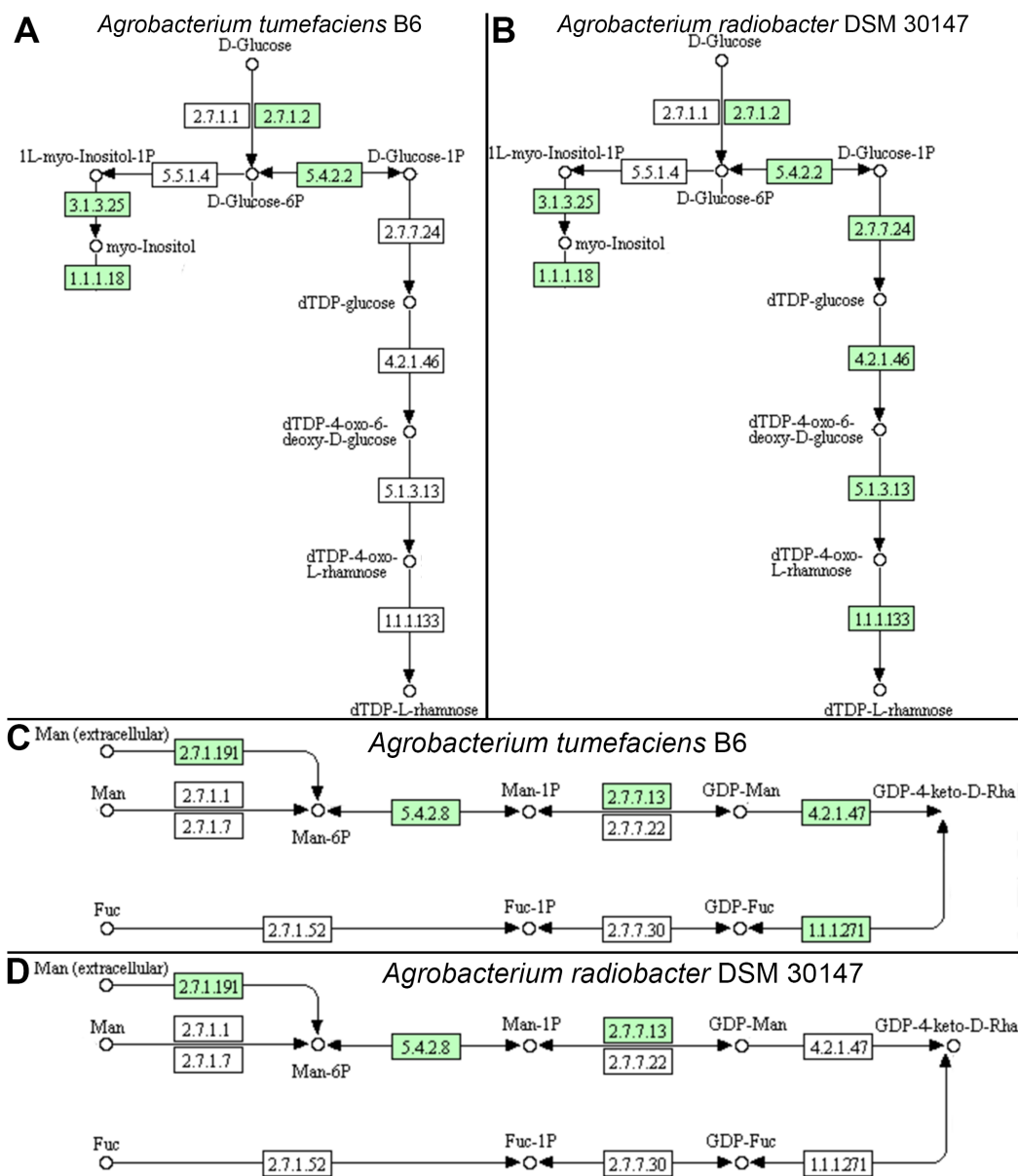


B



**Figure 4** Genomic divergence among genomespecies 4 strains. (A) Unrooted maximum likelihood tree constructed based on the core genome alignment. Branch length and node labels indicate number of substitutions per site and FastTree2 SH-like support values, respectively. Putative subclades were colored blue, red and purple (B) Distribution of accessory (non-core) gene clusters among strains determined with Roary and plotted with the perl script roary2svg.pl (<https://github.com/sanger-pathogens/Roary/blob/master/contrib/roary2svg/roary2svg.pl>). A total of 7,906 accessory gene clusters were identified by Roary and the number of accessory genes presence in each genome are shown in the most right column. Vertical gray lines/bars along the plot indicate presence of accessory gene. Asterisks indicate genomes sequenced in this study. Full-size DOI: 10.7717/peerj.6366/fig-4

their different sub-cluster placement in the pairwise ANI heatmap (Fig. 2). Furthermore, strains from both subclades could be broadly differentiated by the set of core accessory genes that they harbor (Fig. 4B). Therefore, even though *Agrobacterium radiobacter* does not harbor a Ti plasmid, it cannot be considered as a non-tumorigenic strain of *Agrobacterium tumefaciens* given multiple lines of evidence indicating its substantial genomic divergence from *Agrobacterium tumefaciens*.



**Figure 5** KEGG pathway of nucleotide sugar metabolism associated with *Agrobacterium* lipopolysaccharide synthesis. (A & B) genomic potential of *A. tumefaciens* B6 and *A. radiobacter* DSM 30147, respectively, in the biosynthesis of dTDP-L-rhamnose. (C & D) genomic potential of *A. tumefaciens* B6 and *A. radiobacter* DSM 30147, respectively, in the biosynthesis of GDP-L-Fucose. Numbers in boxes indicate Enzyme Commission numbers. White and green boxes indicate absence and presence of the corresponding enzymes, respectively, based on GhostKoala annotation (Kanehisa, Sato & Morishima, 2016b). [Full-size !\[\]\(ba1b80118482ccef74a5d718ca4d7242\_img.jpg\) DOI: 10.7717/peerj.6366/fig-5](https://doi.org/10.7717/peerj.6366/fig-5)

## *Agrobacterium* genomospecies 4 strains differ in their genomic potential for nucleotide sugar metabolism

Individual comparison of the reconstructed KEGG pathways in *Agrobacterium tumefaciens* (Fig. 5A) and *Agrobacterium radiobacter* (Fig. 5B) revealed stark contrast in the anabolism of dTDP-L-rhamnose which is commonly found in the O-antigen of

**Table 2** Identification of *Agrobacterium* proteins with TIGRFAM domains involved in the biosynthesis of nucleotide sugar.

Assembly ID	Strain	TIGR01479 (EC 5.4.2.8)		TIGR01472 (EC 4.2.1.47)	TIGR01207 (EC 2.7.7.24)	TIGR01181 (EC 4.2.1.46)	TIGR01221 (EC 5.1.3.13)	TIGR01214 (EC 1.1.1.133)
		1st hit	2nd hit					
GCF_900045375	B6	690.2	566.6	589.5				
GCF_001541315	B6	690.2	566.6	589.5				
GCF_001692245	B140/95	690.2	566.6	589.5				
GCF_900011755	Kerr14	691.3	690.2	428.6*				
GCF_001541305	NCPPB3001	690.2			494.6	488.5	215.4	331.5
GCF_002008215	LMG140	690.2			494.6	488.5	215.4	331.5
GCF_900012605	CFBP5621	689.3			494.6	489.5	215.4	331.5
GCF_002591665	186	689.3			494.6	488.5	215.4	331.8
GCF_003031125	LAD9	688.5			494.4	487.9	215.4	329.9
GCF_000233975	CCNWGS	644.8			494.6	487.5	215.4	331.8
GCF_002179795	LMG215	690.2						
GCF_000384555	224MFTsu31	644.8						
GCF_000482285	UNC420CL41Cvi	644.8						
GCF_900188475	719_389	687.5						

**Notes:**

Numbers indicate bit scores calculated based on protein alignment to the model with higher scores indicating stronger and more significant hits.

\* Formed a separate protein cluster from the rest of genomospecies 4 GDP-mannose-4,6-dehydratase orthologs (<70% pairwise protein identity).

lipopolysaccharide (LPS) in gram-negative bacteria. Surprisingly, the entire enzyme set required for the generation of dTDP-L-rhamnose from D-glucose-phosphate (Table 2) is absent in *Agrobacterium tumefaciens* B6, suggesting that this common nucleotide sugar may be absent from the LPS O-antigen of strain B6. A manual inspection of the accessory genes uniquely shared by *Agrobacterium tumefaciens* strains B6 and B140/95 identified a homolog cluster containing GDP-L-fucose synthase (EC 1.1.1.271) that is involved in the enzymatic production of GDP-L-fucose from GDP-4-dehydro-6-deoxy-D-mannose and NADH (Table 2; Fig. 5C). As expected, the genes coding for this enzyme and GDP-mannose 4,6-dehydratase involved in the conversion of GDP-alpha-D-mannose to GDP-4-dehydro-6-deoxy-D-mannose, are absent in the *Agrobacterium radiobacter* NCPPB3001 genome (Fig. 5D). Intriguingly, HMMsearch scan revealed the presence of two protein hits to the TIGR01479 HMM profile in *Agrobacterium tumefaciens* B6 that corresponds to D-mannose 1,6-phosphomutase (EC 5.4.2.8) required for the synthesis of D-mannose 6-phosphate. In addition to strain B6, its close relative, strain B140/95, and a more distantly related strain Kerr14 also harbor two copies of this gene. However, one of the D-mannose 1,6-phosphomutases in strain Kerr14 is more divergent with a lower TIGRFAM HMM sequence score (Table 2). Furthermore, it exhibits less than 70% protein identity to the *Agrobacterium tumefaciens* B6 and B140/95 homologs, forming a private protein cluster in the pan-genome (data not shown).

## DISCUSSION

We re-sequenced the genome of *Agrobacterium radiobacter* type strain using strain directly obtained from NCPPB. The assembled *Agrobacterium radiobacter* genome

reported in this study exhibits assembly statistics that are consistent with a high-quality draft genome such as high genome completeness and contiguity, near-zero contamination/duplication and comparable genome size to other closely related strains ([Gan, Lee & Savka, 2018](#); [Parks et al., 2015](#)). Furthermore, given the improved contiguity and dramatic reduction in the number of contigs of this newly assembled draft genome, we recommend using this genome in place of the previously published draft genome for future *Agrobacterium* comparative studies.

The distinct separation of *Agrobacterium* genomospecies 4 and 7 at 95% ANI cutoff corroborates with the previously established “genomic yardstick” for species differentiation ([Konstantinidis & Tiedje, 2005](#); [Richter & Rosselló-Móra, 2009](#)). Using this percentage cutoff, the ANI approach has been successfully used to provide a near “black-and-white” pattern of species separation in even some of the most diverse bacterial genera such as *Pseudomonas*, *Arcobacter* and *Stenotrophomonas* ([Pérez-Cataluña et al., 2018](#); [Tran, Savka & Gan, 2017](#); [Vinuesa, Ochoa-Sánchez & Contreras-Moreira, 2018](#)). Given the increasing evidence highlighting the robustness and reliability of the ANI approach in species delineation, the pairwise ANI between *Agrobacterium tumefaciens* and *Agrobacterium radiobacter* type strains that is at least 2.5% higher than the 95% cutoff value is rigorous evidence that they belong to the same genomospecies, effectively serving as the final nail in the coffin for the decade-long debate on their taxonomic status. The amalgamation of *Agrobacterium radiobacter* and *Agrobacterium tumefaciens* into a single species have been repeatedly suggested in the past few years but was complicated by the special status of *Agrobacterium tumefaciens* as the type species of the genus *Agrobacterium* despite the priority that *Agrobacterium radiobacter* has over *Agrobacterium tumefaciens* as it was isolated and described 3 years before *Agrobacterium tumefaciens* ([Young et al., 2001, 2003](#)). Despite sharing numerous morphological and biochemical features, differences in genomic features such as pairwise ANI, phylogenomic clustering and core accessory gene contents do exist among members in *Agrobacterium* genomospecies 4 that can facilitate the identification of genotypic and phenotypic variants to accurately delimit sub-species relationships in the future ([Brenner, Staley & Krieg, 2000](#); [Jezbera et al., 2011](#); [Meier-Kolthoff et al., 2014](#); [Tan et al., 2013](#)).

To date the LPS for both type strains have been determined ([De Castro et al., 2002, 2004](#)). In stark contrast to *Agrobacterium radiobacter*, the *Agrobacterium tumefaciens* LPS consists of D-arabinose and L-fucose that have yet been reported to date in another members of the genus *Agrobacterium* ([De Castro et al., 2002](#)). The presence of the L configuration of fucose is considered to be rare even among plant pathogenic bacteria but may be associated with the ability of *Agrobacterium tumefaciens* to colonize or bind to wounded plant cell ([Lippincott, Whatley & Lippincott, 1977](#); [Whatley et al., 1976](#); [Whatley & Spiess, 1977](#)). It has been previously shown that the LPS of *Agrobacterium tumefaciens* but not *Agrobacterium radiobacter* can bind to the plant cells thus providing protection against subsequent infection by pathogenic strains ([Whatley et al., 1976](#)). The presence and absence of nucleotide sugars in the O-chain constituent of LPS in both type strains corroborates with their observed genomic potential in the nucleotide sugar metabolism pathway thus underscoring the utility of comparative genomics in facilitating

the prediction of microbial host range and ecological niche (Klosterman *et al.*, 2011). For example, the absence of L-rhamnose and L-fucose in the LPS of *Agrobacterium tumefaciens* B6 and *Agrobacterium radiobacter* DSM30147, respectively, is consistent with the lack of genes coding for enzymes involved with the particular nucleotide sugar metabolism. Generation of *Agrobacterium tumefaciens* B6 LPS mutant via targeted gene deletion (Kaczmarczyk, Vorholt & Francez-Charlot, 2012) or the classical but more laborious transposon mutagenesis approach followed by characterization of the LPS mutant host-range and phytopathogenicity will be instructive (Gan *et al.*, 2011; Reuhs *et al.*, 2005).

Our current genomic sampling indicates that the Ti plasmid appears to be restricted to the *Agrobacterium tumefaciens* subclade. The maintenance of the Ti plasmid is metabolically taxing given its large size (Barker *et al.*, 1983; Glick, 1995). Even if the Ti plasmid was conjugally transfer, for example, to *Agrobacterium radiobacter*, the inability of *Agrobacterium radiobacter* to colonize plant host as evidenced by its LPS incompatibility will not confer an advantage to the new plasmid host in a natural environment (Thomashow *et al.*, 1980). Furthermore, in the absence of high density Acyl-homoserine lactone (AHL) signals which is required to trigger Ti plasmid conjugation (Fuqua & Winans, 1994; Pappas, 2008; Zhang, Wang & Zhang, 2002), the newly acquired Ti plasmid in *Agrobacterium radiobacter* may be cured in its natural soil habitat after a few generations. Although the spontaneous transfer of the Ti plasmid from tumorigenic *Agrobacterium tumefaciens* to *Agrobacterium radiobacter* K84 has been reported previously, strain K84 was re-classified based on a recent core gene analysis to *Rhizobium rhizogenes* K84 (Velázquez *et al.*, 2010; Vicedo *et al.*, 1996), reiterating the pervasive taxonomic inconsistency within the genus *Agrobacterium* that may have confound previous biological interpretations (De Ley *et al.*, 1966; Lindström *et al.*, 1995; Young, 2008). Given that a large majority of *Agrobacterium* genetics was performed during the pre-NGS era (Gan & Savka, 2018), it remains unknown as to how many *Agrobacterium tumefaciens* and *Agrobacterium radiobacter* strains have been molecularly misclassified due to their high genomic relatedness.

The inability to accurately identify plasmid and chromosomal-derived contigs among the draft genomes means that some of the core accessory genes among tumorigenic strains may be plasmid-derived and should be treated with caution as the low-copy-number Ti-plasmid is prone to curing in the absence of AHL signals. Despite the value of complete genome assembly in enabling the accurate partitioning of plasmid and chromosomal genomic region (Arredondo-Alonso *et al.*, 2017), the representation of complete *Agrobacterium* genomes in current database is still very low as a majority of the genomes were assembled from short Illumina reads that cannot effectively span repetitive region (Wibberg *et al.*, 2011; Wood *et al.*, 2001). Furthermore, most *Agrobacterium* strains harbor multiple large plasmids that further complicate short-read-only assembly graph (Kado & Liu, 1981; Lowe *et al.*, 2009; Shao *et al.*, 2018). Given the currently available genomic resources for *Agrobacterium*, defining subspecies within the *Agrobacterium* genomospecies 4 based on the identification of lineage-specific gene set (Moldovan & Gelfand, 2018) will be challenging. However, we anticipate that the advent of

high throughput long-read sequencing that can span large repetitive region in recent years is likely going to overcome this limitation allowing a more accurate depiction of microbial pangenome (*Gan et al., 2012; Gan, Lee & Austin, 2017; Schmid et al., 2018a, 2018b*). Future hybrid genome assemblies (Illumina and Nanopore/PacBio reads) of members from genomospecies 4 with comprehensive metadata and reliable phenotypic information, will be instructive.

## CONCLUSIONS

Despite belonging to the same genomospecies, *Agrobacterium tumefaciens* and *Agrobacterium radiobacter* are by no means clonal at the chromosomal level and instead demonstrate sufficient genomic characters that qualify their separation into two sub-species. In addition, the difference in the LPS profile among two type strains will have implications to host specificity leading to geographical separation. In the spirit of preserving the naming of both species but at the same time respecting the taxonomic jurisdiction for strain priority, we propose *Agrobacterium tumefaciens* to be reclassified as *Agrobacterium radiobacter subsp. tumefaciens* and for *Agrobacterium radiobacter* to retain its species status with the proposed name of *Agrobacterium radiobacter subsp. radiobacter*.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

Michael A. Savka and Han Ming Gan received support from the College of Science and the Thomas H. Gosnell School of Life Sciences at Rochester Institute of Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
College of Science.  
School of Life Sciences at Rochester Institute of Technology.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Han Ming Gan conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Melvin V.L. Lee performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Michael A. Savka conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Raw sequencing data and whole genome assembly for strains B6 and NCPPB3001 reported in this study are linked to the NCBI Bioproject IDs [PRJNA300485](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA300485) and [PRJNA300611](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA300611), respectively.

## Data Availability

The following information was supplied regarding data availability:

LMVK00000000.1, ASM154131v1: [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_001541315.1](https://www.ncbi.nlm.nih.gov/assembly/GCF_001541315.1);

LMVJ00000000.1, ASM154130v1: [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_001541305.1](https://www.ncbi.nlm.nih.gov/assembly/GCF_001541305.1);

Code and data are available at Han Ming Gan. (2018). Dataset for “Improved genome of *Agrobacterium radiobacter* type strain provides new taxonomic insight into *Agrobacterium* genomospecies 4” [Data set]. Zenodo. DOI [10.5281/zenodo.1489356](https://doi.org/10.5281/zenodo.1489356).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.6366#supplemental-information>.

## REFERENCES

- Alikhan N-F, Petty NK, Zakour NLB, Beatson SA. 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**(1):1 DOI [10.1186/1471-2164-12-402](https://doi.org/10.1186/1471-2164-12-402).
- Allnutt T, Yan CZY, Crowley TM, Gan HM. 2018. Commentary: genome sequence of *Vibrio parahaemolyticus* VP152 strain isolated from *Penaeus indicus* in malaysia. *Frontiers in microbiology* **9**:865 DOI [10.3389/fmicb.2018.00865](https://doi.org/10.3389/fmicb.2018.00865).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3):403–410 DOI [10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2).
- Arredondo-Alonso S, Willems RJ, Van Schaik W, Schürch AC. 2017. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics* **3**(10):e000128 DOI [10.1099/mgen.0.000128](https://doi.org/10.1099/mgen.0.000128).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5):455–477 DOI [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- Barker R, Idler K, Thompson D, Kemp J. 1983. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. *Plant Molecular Biology* **2**(6):335–350 DOI [10.1007/bf01578595](https://doi.org/10.1007/bf01578595).
- Beijerinck M, Van Delden A. 1902. On a colourless bacterium, whose carbon food comes from the atmosphere. *Koninklijke Nederlandse Akademie Van Wetenschappen Proceedings Series B Physical Sciences* **5**:398–413.
- Bobay L-M, Ochman H. 2017. Biological species are universal across life’s domains. *Genome Biology and Evolution* **9**(3):491–501 DOI [10.1093/gbe/evx026](https://doi.org/10.1093/gbe/evx026).



- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Bourras S, Rouxel T, Meyer M. 2015.** *Agrobacterium tumefaciens* gene transfer: how a plant pathogen hacks the nuclei of plant and nonplant organisms. *Phytopathology* **105**(10):1288–1301 DOI [10.1094/phyto-12-14-0380-rvw](https://doi.org/10.1094/phyto-12-14-0380-rvw).
- Bouzar H, Jones JB. 2001.** *Agrobacterium larrymoorei* sp. nov., a pathogen isolated from aerial tumours of *Ficus benjamina*. *International Journal of Systematic and Evolutionary Microbiology* **51**(3):1023–1026 DOI [10.1099/00207713-51-3-1023](https://doi.org/10.1099/00207713-51-3-1023).
- Brenner D, Staley J, Krieg N. 2000.** Classification of procaryotic organisms and the concept of bacterial speciation. In: Boone DR, Castenholz RW, Garrity GM, eds. *Bergey's Manual of Systematic Bacteriology*. Second Edition. Vol. 1. New York: Springer Verlag, 27–38 DOI [10.1002/9781118960608.bm00006](https://doi.org/10.1002/9781118960608.bm00006).
- Conn HJ. 1942.** Validity of the genus *Alcaligenes*. *Journal of Bacteriology* **44**:353–360.
- Coutinho F, Tschoeke DA, Thompson F, Thompson C. 2016.** Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. *PeerJ* **4**:e1522 DOI [10.7717/peerj.1522](https://doi.org/10.7717/peerj.1522).
- De Castro C, Bedini E, Garozzo D, Sturiale L, Parrilli M. 2004.** Structural determination of the O-chain moieties of the lipopolysaccharide fraction from *Agrobacterium radiobacter* DSM 30147. *European Journal of Organic Chemistry* **2004**:3842–3849 DOI [10.1002/ejoc.200400238](https://doi.org/10.1002/ejoc.200400238).
- De Castro C, De Castro O, Molinaro A, Parrilli M. 2002.** Structural determination of the O-chain polysaccharide from *Agrobacterium tumefaciens*, strain DSM 30205. *European Journal of Biochemistry* **269**(12):2885–2888 DOI [10.1046/j.1432-1033.2002.02955.x](https://doi.org/10.1046/j.1432-1033.2002.02955.x).
- De Ley J, Bernaerts M, Rassel A, Guilmot J. 1966.** Approach to an improved taxonomy of the genus *Agrobacterium*. *Journal of General Microbiology* **43**(1):7–17 DOI [10.1099/00221287-43-1-7](https://doi.org/10.1099/00221287-43-1-7).
- Farrand SK, Van Berkum PB, Oger P. 2003.** *Agrobacterium* is a definable genus of the family Rhizobiaceae. *International Journal of Systematic and Evolutionary Microbiology* **53**(5):1681–1687 DOI [10.1099/ijs.0.02445-0](https://doi.org/10.1099/ijs.0.02445-0).
- Fuqua WC, Winans SC. 1994.** A LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. *Journal of Bacteriology* **176**(10):2796–2806 DOI [10.1128/jb.176.10.2796-2806.1994](https://doi.org/10.1128/jb.176.10.2796-2806.1994).
- Gan HM, Chew TH, Tay Y-L, Lye SF, Yahya A. 2012.** Genome sequence of *Hydrogenophaga* sp. strain PBC, a 4-aminobenzenesulfonate-degrading bacterium. *Journal of Bacteriology* **194**(17):4759–4760 DOI [10.1128/jb.00990-12](https://doi.org/10.1128/jb.00990-12).
- Gan HM, Gan HY, Ahmad NH, Aziz NA, Hudson AO, Savka MA. 2015.** Whole genome sequencing and analysis reveal insights into the genetic structure, diversity and evolutionary relatedness of luxI and luxR homologs in bacteria belonging to the Sphingomonadaceae family. *Frontiers in Cellular and Infection Microbiology* **4**:188 DOI [10.3389/fcimb.2014.00188](https://doi.org/10.3389/fcimb.2014.00188).
- Gan HY, Gan HM, Savka MA, Triassi AJ, Wheatley MS, Smart LB, Fabio ES, Hudson AO. 2014.** Whole-genome sequences of 13 endophytic bacteria isolated from shrub willow (*Salix*) grown in Geneva, New York. *Genome Announcements* **2**(3):e00288-14 DOI [10.1128/genomeA.00288-14](https://doi.org/10.1128/genomeA.00288-14).
- Gan HM, Ibrahim Z, Shahir S, Yahya A. 2011.** Identification of genes involved in the 4-aminobenzenesulfonate degradation pathway of *Hydrogenophaga* sp. PBC via transposon mutagenesis. *FEMS Microbiology Letters* **318**(2):108–114 DOI [10.1111/j.1574-6968.2011.02245.x](https://doi.org/10.1111/j.1574-6968.2011.02245.x).

- Gan HM, Lee YP, Austin CM. 2017.** Nanopore long-read guided complete genome assembly of *Hydrogenophaga intermedia*, and genomic insights into 4-aminobenzenesulfonate, p-aminobenzoic acid and hydrogen metabolism in the genus *Hydrogenophaga*. *Frontiers in Microbiology* **8**:1880 DOI [10.3389/fmicb.2017.01880](https://doi.org/10.3389/fmicb.2017.01880).
- Gan HM, Lee MVJ, Savka MA. 2018.** High-quality draft genome sequence of the type strain of *Allorhizobium vitis*, the primary causal agent of grapevine crown gall. *Microbiology Resource Announcements* **7**(9):e01045-18 DOI [10.1128/mra.01045-18](https://doi.org/10.1128/mra.01045-18).
- Gan HM, Savka MA. 2018.** One more decade of *Agrobacterium* Taxonomy. *Current Topics in Microbiology and Immunology* **418**:1–14 DOI [10.1007/82\\_2018\\_81](https://doi.org/10.1007/82_2018_81).
- Gan HM, Tan MH, Austin CM. 2016a.** The complete mitogenome of the red claw crayfish *Cherax quadricarinatus* (Von Martens, 1868)(Crustacea: Decapoda: Parastacidae). *Mitochondrial DNA Part A* **27**(1):385–386 DOI [10.3109/19401736.2014.895997](https://doi.org/10.3109/19401736.2014.895997).
- Gan HM, Tan MH, Eprilurahman R, Austin CM. 2016b.** The complete mitogenome of *Cherax monticola* (Crustacea: Decapoda: Parastacidae), a large highland crayfish from New Guinea. *Mitochondrial DNA Part A* **27**:337–338 DOI [10.3109/19401736.2014.892105](https://doi.org/10.3109/19401736.2014.892105).
- Gaunt M, Turner S, Rigottier-Gois L, Lloyd-Macgilp S, Young J. 2001.** Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *International Journal of Systematic and Evolutionary Microbiology* **51**(6):2037–2048 DOI [10.1099/00207713-51-6-2037](https://doi.org/10.1099/00207713-51-6-2037).
- Glick BR. 1995.** Metabolic load and heterologous gene expression. *Biotechnology Advances* **13**(2):247–261 DOI [10.1016/0734-9750\(95\)00004-a](https://doi.org/10.1016/0734-9750(95)00004-a).
- Gordon JE, Christie PJ. 2014.** The *Agrobacterium* Ti plasmids. *Microbiology Spectrum* **2**(6):PLAS-0010-2013 DOI [10.1128/microbiolspec.PLAS-0010-2013](https://doi.org/10.1128/microbiolspec.PLAS-0010-2013).
- Haft DH, Selengut JD, White O. 2003.** The TIGRFAMs database of protein families. *Nucleic Acids Research* **31**(1):371–373 DOI [10.1093/nar/gkg128](https://doi.org/10.1093/nar/gkg128).
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010.** Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**(1):119 DOI [10.1186/1471-2105-11-119](https://doi.org/10.1186/1471-2105-11-119).
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018.** High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**:5114 DOI [10.1038/s41467-018-07641-9](https://doi.org/10.1038/s41467-018-07641-9).
- Jeong H, Pan J-G, Park S-H. 2016.** Contamination as a major factor in poor Illumina assembly of microbial isolate genomes. *bioRxiv preprint* DOI [10.1101/081885](https://doi.org/10.1101/081885).
- Jezbera J, Jezberová J, Brandt U, Hahn MW. 2011.** Ubiquity of *Polynucleobacter necessarius* subspecies *asymbioticus* results from ecological diversification. *Environmental Microbiology* **13**(4):922–931 DOI [10.1111/j.1462-2920.2010.02396.x](https://doi.org/10.1111/j.1462-2920.2010.02396.x).
- Johnson LS, Eddy SR, Portugaly E. 2010.** Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**(1):431 DOI [10.1186/1471-2105-11-431](https://doi.org/10.1186/1471-2105-11-431).
- Kaczmarczyk A, Vorholt JA, Francez-Charlot A. 2012.** Markerless gene deletion system for sphingomonads. *Applied and Environmental Microbiology* **78**(10):3774–3777 DOI [10.1128/AEM.07347-11](https://doi.org/10.1128/AEM.07347-11).
- Kado CI, Liu ST. 1981.** Rapid procedure for detection and isolation of large and small plasmids. *Journal of Bacteriology* **145**(3):1365–1373.
- Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermini LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**(6):587–589 DOI [10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285).

- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016a. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**(D1):D457–D462 DOI [10.1093/nar/gkv1070](https://doi.org/10.1093/nar/gkv1070).
- Kanehisa M, Sato Y, Morishima K. 2016b. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology* **428**(4):726–731 DOI [10.1016/j.jmb.2015.11.006](https://doi.org/10.1016/j.jmb.2015.11.006).
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**(4):772–780 DOI [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Keane PJ, Kerr A, New PB. 1970. Crown gall of stone fruit II. Identification and nomenclature of *Agrobacterium* isolates. *Australian Journal of Biological Sciences* **23**(3):585–596 DOI [10.1071/bi9700585](https://doi.org/10.1071/bi9700585).
- Kerr A, Panagopoulos CG. 1977. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. *Journal of Phytopathology* **90**(2):172–179 DOI [10.1111/j.1439-0434.1977.tb03233.x](https://doi.org/10.1111/j.1439-0434.1977.tb03233.x).
- Kim K, Gan HM. 2017. A glimpse into the genetic basis of symbiosis between *Hydrogenophaga* and their helper strains in the biodegradation of 4-aminobenzenesulfonate. *Journal of Genomics* **5**:77–82 DOI [10.7150/jgen.20216](https://doi.org/10.7150/jgen.20216).
- Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, De Jonge R, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma L-J. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLOS Pathogens* **7**(7):e1002137 DOI [10.1371/journal.ppat.1002137](https://doi.org/10.1371/journal.ppat.1002137).
- Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**(7):2567–2572 DOI [10.1073/pnas.0409727102](https://doi.org/10.1073/pnas.0409727102).
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**(13):1658–1659 DOI [10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158).
- Lindström K, Van Berkum P, Gillis M, Martinez E, Novikova N, Jarvis B. 1995. Report from the roundtable on Rhizobium taxonomy. In: Tikhonovich IA, Provorov NA, Romanov VI, Newton WE, eds. *Nitrogen Fixation. Fundamentals and Applications*. Kluwer, Dordrecht: Springer, 807–810.
- Lippincott BB, Whatley MH, Lippincott JA. 1977. Tumor induction by *Agrobacterium* involves attachment of the bacterium to a site on the host plant cell wall. *Plant Physiology* **59**(3):388–390 DOI [10.1104/pp.59.3.388](https://doi.org/10.1104/pp.59.3.388).
- Lowe N, Gan HM, Chakravarty V, Scott R, Szegedi E, Burr TJ, Savka MA. 2009. Quorum-sensing signal production by *Agrobacterium vitis* strains and their tumor-inducing and tartrate-catabolic plasmids. *FEMS Microbiology Letters* **296**(1):102–109 DOI [10.1111/j.1574-6968.2009.01627.x](https://doi.org/10.1111/j.1574-6968.2009.01627.x).
- Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy TBK, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk H-P. 2014. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences* **9**(1):2 DOI [10.1186/1944-3277-9-2](https://doi.org/10.1186/1944-3277-9-2).

- Moldovan MA, Gelfand MS. 2018.** Pangenomic definition of prokaryotic species and the phylogenetic structure of *Prochlorococcus* spp. *Frontiers in Microbiology* **9**:428 DOI [10.3389/fmicb.2018.00428](https://doi.org/10.3389/fmicb.2018.00428).
- Mousavi SA, Willems A, Nesme X, De Lajudie P, Lindstrom K. 2015.** Revised phylogeny of Rhizobiaceae: proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations. *Systematic and Applied Microbiology* **38**(2):84–90 DOI [10.1016/j.syapm.2014.12.003](https://doi.org/10.1016/j.syapm.2014.12.003).
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2014.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**(1):268–274 DOI [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).
- Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015.** Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**(22):3691–3693 DOI [10.1093/bioinformatics/btv421](https://doi.org/10.1093/bioinformatics/btv421).
- Panagopoulos C, Psallidas P, Alivizatos A. 1978.** Studies on biotype 3 of *Agrobacterium radiobacter* var. *tumefaciens*. In: *Proceedings of the 4th International Conference on Plant Pathogenic Bacteria Vol 1*. Institut National de la Recherche Agronomique, Anger, France: Sta. Path. Veg. Phytobact, 221–228.
- Pappas KM. 2008.** Cell–cell signaling and the *Agrobacterium tumefaciens* Ti plasmid copy number fluctuations. *Plasmid* **60**(2):89–107 DOI [10.1016/j.plasmid.2008.05.003](https://doi.org/10.1016/j.plasmid.2008.05.003).
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015.** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**(7):1043–1055 DOI [10.1101/gr.186072.114](https://doi.org/10.1101/gr.186072.114).
- Pérez-Cataluña A, Salas-Massó N, Dieguez ABL, Balboa S, Lema A, Romalde JL, Figueras MJ. 2018.** Revisiting the taxonomy of the genus *Arcobacter*: getting order from the chaos. *Frontiers in Microbiology* **9**:2077 DOI [10.3389/fmicb.2018.02077](https://doi.org/10.3389/fmicb.2018.02077).
- Price MN, Dehal PS, Arkin AP. 2010.** FastTree 2—approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**:e9490.
- Ramírez-Bahena MH, Vial L, Lassalle F, Diel B, Chapulliot D, Daubin V, Nesme X, Muller D. 2014.** Single acquisition of protelomerase gave rise to speciation of a large and diverse clade within the *Agrobacterium/Rhizobium* supercluster characterized by the presence of a linear chromid. *Molecular Phylogenetics and Evolution* **73**:202–207 DOI [10.1016/j.ympev.2014.01.005](https://doi.org/10.1016/j.ympev.2014.01.005).
- Reuhs BL, Relić B, Forsberg LS, Marie C, Ojanen-Reuhs T, Stephens SB, Wong C-H, Jabbouri S, Broughton WJ. 2005.** Structural characterization of a flavonoid-inducible *Pseudomonas aeruginosa* A-Band-Like O antigen of *Rhizobium* sp. strain NGR234, required for the formation of nitrogen-fixing nodules. *Journal of Bacteriology* **187**(18):6479–6487 DOI [10.1128/jb.187.18.6479-6487.2005](https://doi.org/10.1128/jb.187.18.6479-6487.2005).
- Richter M, Rosselló-Móra R. 2009.** Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**(45):19126–19131 DOI [10.1073/pnas.0906412106](https://doi.org/10.1073/pnas.0906412106).
- Riker A, Banfield W, Wright W, Keitt G, Sagen HE. 1930.** Studies on infectious hairy root of nursery apple trees. *Journal of Agricultural Research* **41**(7):507–540.
- Sawada H, Ieki H, Oyaizu H, Matsumoto S. 1993.** Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *International Journal of Systematic Bacteriology* **43**(4):694–702 DOI [10.1099/00207713-43-4-694](https://doi.org/10.1099/00207713-43-4-694).

- Schmid M, Frei D, Patrignani A, Schlapbach R, Frey JE, Remus-Emsermann MNP, Ahrens CH. 2018a. Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. *Nucleic Acids Research* **46**:8953–8965 DOI [10.1093/nar/gky726](https://doi.org/10.1093/nar/gky726).
- Schmid M, Muri J, Melidis D, Varadarajan AR, Somerville V, Wicki A, Moser A, Bourqui M, Wenzel C, Eugster-Meier E, Frey JE, Irmeler S, Ahrens CH. 2018b. Comparative genomics of completely sequenced *Lactobacillus helveticus* genomes provides insights into strain-specific genes and resolves metagenomics data down to the strain level. *Frontiers in Microbiology* **9**:63 DOI [10.3389/fmicb.2018.00063](https://doi.org/10.3389/fmicb.2018.00063).
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**(14):2068–2069 DOI [10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153).
- Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications* **4**(1):2304 DOI [10.1038/ncomms3304](https://doi.org/10.1038/ncomms3304).
- Shams M, Vial L, Chapulliot D, Nesme X, Lavire C. 2013. Rapid and accurate species and genomic species identification and exhaustive population diversity assessment of *Agrobacterium* spp. using recA-based PCR. *Systematic and Applied Microbiology* **36**(5):351–358 DOI [10.1016/j.syapm.2013.03.002](https://doi.org/10.1016/j.syapm.2013.03.002).
- Shao S, Zhang X, Van Heusden GPH, Hooykaas PJ. 2018. Complete sequence of the tumor-inducing plasmid pTiChry5 from the hypervirulent *Agrobacterium tumefaciens* strain Chry5. *Plasmid* **96–97**:1–6 DOI [10.1016/j.plasmid.2018.02.001](https://doi.org/10.1016/j.plasmid.2018.02.001).
- Smith EF, Townsend CO. 1907. A plant-tumor of bacterial origin. *Science* **25**(643):671–673 DOI [10.1126/science.25.643.671](https://doi.org/10.1126/science.25.643.671).
- Sokolov EP. 2000. An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *Journal of Molluscan Studies* **66**(4):573–575 DOI [10.1093/mollus/66.4.573](https://doi.org/10.1093/mollus/66.4.573).
- Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC, Nesme X, Rosselló-Mora R, Swings J, Trüper HG. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology* **52**(3):1043–1047.
- Starr M, Weiss J. 1943. Growth of phytopathogenic bacteria in a synthetic asparagin medium. *Phytopathology* **33**:314–318.
- Süle S. 1978. Biotypes of *Agrobacterium tumefaciens* in Hungary. *Journal of Applied Bacteriology* **44**(2):207–213 DOI [10.1111/j.1365-2672.1978.tb00792.x](https://doi.org/10.1111/j.1365-2672.1978.tb00792.x).
- Tan MH, Gan HM, Schultz MB, Austin CM. 2015. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. *Molecular Phylogenetics and Evolution* **85**:180–188 DOI [10.1016/j.ympev.2015.02.009](https://doi.org/10.1016/j.ympev.2015.02.009).
- Tan JL, Khang TF, Ngeow YF, Choo SW. 2013. A phylogenomic approach to bacterial subspecies classification: proof of concept in *Mycobacterium abscessus*. *BMC Genomics* **14**(1):879 DOI [10.1186/1471-2164-14-879](https://doi.org/10.1186/1471-2164-14-879).
- Thomashow M, Panagopoulos C, Gordon M, Nester E. 1980. Host range of *Agrobacterium tumefaciens* is determined by the Ti plasmid. *Nature* **283**(5749):794–796 DOI [10.1038/283794a0](https://doi.org/10.1038/283794a0).
- Tindall B. 2014. *Agrobacterium radiobacter* (Beijerinck and van Delden 1902) Conn 1942 has priority over *Agrobacterium tumefaciens* (Smith and Townsend 1907) Conn 1942 when the two are treated as members of the same species based on the principle of priority and

- Rule 23a, Note 1 as applied to the corresponding specific epithets. Opinion 94: judicial commission of the international committee on systematics of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **64**(Pt 10):3590–3592 DOI [10.1099/ijls.0.069203-0](https://doi.org/10.1099/ijls.0.069203-0).
- Tran PN, Savka MA, Gan HM. 2017.** In-silico taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus *Pseudomonas*. *Frontiers in Microbiology* **8**:1296 DOI [10.3389/fmicb.2017.01296](https://doi.org/10.3389/fmicb.2017.01296).
- Utturkar SM, Klingeman DM, Hurt RA, Brown SD. 2017.** A case study into microbial genome assembly gap sequences and finishing strategies. *Frontiers in Microbiology* **8**:1272 DOI [10.3389/fmicb.2017.01272](https://doi.org/10.3389/fmicb.2017.01272).
- Velázquez E, Palomo JL, Rivas R, Guerra H, Peix A, Trujillo ME, García-Benavides P, Mateos PF, Wabiko H, Martínez-Molina E. 2010.** Analysis of core genes supports the reclassification of strains *Agrobacterium radiobacter* K84 and *Agrobacterium tumefaciens* AKE10 into the species *Rhizobium rhizogenes*. *Systematic and Applied Microbiology* **33**(5):247–251 DOI [10.1016/j.syapm.2010.04.004](https://doi.org/10.1016/j.syapm.2010.04.004).
- Vicedo B, López MJ, Asíns MJ, López MM. 1996.** Spontaneous transfer of the Ti plasmid of *Agrobacterium tumefaciens* and the nopaline catabolism plasmid of *A. radiobacter* strain K84 in crown gall tissue. *Phytopathology* **86**(5):528–534 DOI [10.1094/phyto-86-528](https://doi.org/10.1094/phyto-86-528).
- Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B. 2018.** GET\_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Frontiers in Microbiology* **9**:771 DOI [10.3389/fmicb.2018.00771](https://doi.org/10.3389/fmicb.2018.00771).
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017.** BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**(3):543–548 DOI [10.1093/molbev/msx319](https://doi.org/10.1093/molbev/msx319).
- Whatley M, Bodwin J, Lippincott B, Lippincott J. 1976.** Role of *Agrobacterium* cell envelope lipopolysaccharide in infection site attachment. *Infection and Immunity* **13**:1080–1083.
- Whatley MH, Spiess LD. 1977.** Role of bacterial lipopolysaccharide in attachment of *Agrobacterium* to moss. *Plant Physiology* **60**:765–766.
- Wibberg D, Blom J, Jaenicke S, Kollin F, Rupp O, Scharf B, Schneiker-Bekel S, Sczcepanowski R, Goesmann A, Setubal JC, Schmitt R, Pühler A, Schlüter A. 2011.** Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *Journal of Biotechnology* **155**(1):50–62 DOI [10.1016/j.jbiotec.2011.01.010](https://doi.org/10.1016/j.jbiotec.2011.01.010).
- Wong YM, Juan JC, Gan HM, Austin CM. 2014.** Draft genome sequence of *Clostridium perfringens* strain JJC, a highly efficient hydrogen producer isolated from landfill leachate sludge. *Genome Announcements* **2**(2):e00064-14 DOI [10.1128/genomeA.00064-14](https://doi.org/10.1128/genomeA.00064-14).
- Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF. 2001.** The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**(5550):2317–2323 DOI [10.1126/science.1066804](https://doi.org/10.1126/science.1066804).
- Young JM. 2008.** *Agrobacterium*—taxonomy of plant-pathogenic *Rhizobium* species. *Agrobacterium: From Biology to Biotechnology*. New York: Springer, 183–220.
- Young J, Kuykendall L, Martinez-Romero E, Kerr A, Sawada H. 2001.** A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new

combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *International Journal of Systematic and Evolutionary Microbiology* **51**(1):89–103  
[DOI 10.1099/00207713-51-1-89](https://doi.org/10.1099/00207713-51-1-89).

**Young J, Kuykendall L, Martinez-Romero E, Kerr A, Sawada H. 2003.** Classification and nomenclature of *Agrobacterium* and *Rhizobium*—a reply to Farrand et al. (2003). *International Journal of Systematic and Evolutionary Microbiology* **53**(5):1689–1695  
[DOI 10.1099/ijs.0.02762-0](https://doi.org/10.1099/ijs.0.02762-0).

**Young JM, Pennycook SR, Watson DR. 2006.** Proposal that *Agrobacterium radiobacter* has priority over *Agrobacterium tumefaciens*. Request for an opinion. *International Journal of Systematic and Evolutionary Microbiology* **56**(2):491–493 [DOI 10.1099/ijs.0.64030-0](https://doi.org/10.1099/ijs.0.64030-0).

**Zhang L, Li X, Zhang F, Wang G. 2014.** Genomic analysis of *Agrobacterium radiobacter* DSM 30147T and emended description of *A. radiobacter* (Beijerinck and van Delden 1902) Conn 1942 (Approved Lists 1980) emend. Sawada et al. 1993. *Standards in Genomic Sciences* **9**(3):574–584 [DOI 10.4056/sigs.4688352](https://doi.org/10.4056/sigs.4688352).

**Zhang H-B, Wang L-H, Zhang L-H. 2002.** Genetic control of quorum-sensing signal turnover in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences of the United States of America* **99**(7):4638–4643 [DOI 10.1073/pnas.022056699](https://doi.org/10.1073/pnas.022056699).