Comparative Genomics of Novel Agrobacterium G3 Strains Isolated From the International Space Station and Description of Agrobacterium tomkonis sp. nov.

Singh, Nitin K.; Lavire, Céline; Nesme, Joseph; Vial, Ludovic; Nesme, Xavier; Mason, Christopher E.; Lassalle, Florent; Venkateswaran, Kasthuri

Published in:
Frontiers in Microbiology

DOI:
10.3389/fmicb.2021.765943

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Comparative Genomics of Novel Agrobacterium G3 Strains Isolated From the International Space Station and Description of Agrobacterium tomkonis sp. nov.

Nitin K. Singh\textsuperscript{1†}, Céline Lavire\textsuperscript{2†}, Joseph Nesme\textsuperscript{3†}, Ludovic Vial\textsuperscript{2}, Xavier Nesme\textsuperscript{2}, Christopher E. Mason\textsuperscript{4}, Florent Lassalle\textsuperscript{5*} and Kasthuri Venkateswaran\textsuperscript{1*}

\textsuperscript{1} Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, \textsuperscript{2} CNRS, INRAE, VetAgro Sup, UMR Ecologie Microbienne, Université de Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France, \textsuperscript{3} Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark, \textsuperscript{4} Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, United States, \textsuperscript{5} Parasites and Microbes Programme, Wellcome Sanger Institute, Hinxton, United Kingdom

Strains of Agrobacterium genomospecies 3 (i.e., genomovar G3 of the Agrobacterium tumefaciens species complex) have been previously isolated from diverse environments, including in association with plant roots, with algae, as part of a lignocellulose degrading community, from a hospital environment, as a human opportunistic pathogen, or as reported in this study, from a surface within the International Space Station. Polyphasic taxonomic methods revealed the relationship of Agrobacterium G3 strains to other Agrobacterium spp., which supports the description of a novel species. The G3 strains tested ($n = 9$) were phenotypically distinguishable among the strains from other genomospecies of the genus Agrobacterium. Phylogenetic analyses of the 16S rRNA gene, gyrB gene, multi-locus sequence analysis, and 1,089-gene core-genome gene concatenate concur that tested G3 strains belong to the Agrobacterium genus and they form a clade distinct from other validly described Agrobacterium species. The distinctiveness of this clade was confirmed by average nucleotide identity (ANI) and \textit{in silico} digital DNA–DNA hybridization (dDDH) comparisons between the G3 tested strains and all known Agrobacterium species type strains, since obtained values were considerably below the 95\% (ANI) and 70\% (dDDH) thresholds used for the species delineation. According to the core-genome phylogeny and ANI comparisons, the closest relatives of G3 strains were Agrobacterium sp. strains UGM030330-04 and K599, members of a novel genomospecies we propose to call genomovar G21. Using this polyphasic approach, we characterized the phenotypic and genotypic synapomorphies of Agrobacterium G3, showing it is a \textit{bona fide} bacterial species, well separated from previously named Agrobacterium species or other recognized genomic species. We thus propose the name Agrobacterium tomkonis for this species previously referred to as Agrobacterium genomospecies 3. The type strain of A. tomkonis is IIF1SW-B1\textsuperscript{T} (= LMG
Singh et al. Comparative Genomics of Agrobacterium tomkonis

INTRODUCTION

Agrobacterium species are firstly known for their pathogenic potential on plant, causing either crown gall or hairy root disease, depending on the presence of a tumor-inducing (Ti) or root-inducing (Ri) megaplasmid in their genome; however, most strains lack such plasmids and are not pathogenic. Agrobacterium species are common members of soil communities, and are efficient colonizers of the rhizospheres of a wide variety of plant hosts (Macrae et al., 1988; Matthesse and McMahan, 1998), with which they have a commensal relationship (Sanguin et al., 2006; Dessaux and Faure, 2018). Agrobacterium is a polyphyletic group, which taxonomy, originally aimed at reflecting the pathogenicity status of strains (Smith and Townsend, 1907), was revised multiple times, in the light of growing phenotypic and genotypic evidence, notably showing that the pathogenicity status was not correlated to the diversity of agrobacteria. Based on biochemical testing and DNA-DNA hybridization (DDH), Agrobacterium species were initially classified into three biovars (Keane et al., 1970; Kersters et al., 1973; Ophel and Kerr, 1987). Agrobacterium species belonging to biovar 1, as well as closely related species such as Agrobacterium rubi and Agrobacterium larrymoorei can be recognized by the presence of a linear chromid as the second major molecule of the genome, a synapomorphic trait enabled by the landmark acquisition of the telA gene (Ramirez-Bahena et al., 2014). The strains that belong to biovar 2 and biovar 3 are only distantly related to Agrobacterium biovar 1, which later led to their reclassification into the genera Rhizobium and Allorhizobium, respectively (Young et al., 2001; Mousavi et al., 2015). Further refinement of the taxonomy of Agrobacterium biovar 1, also known as the Agrobacterium tumefaciens species complex, was conducted based on whole-genome information using DDH, leading to the classification of strains into species-level units called genomic species or genomovars (Popoff et al., 1984). This classification framework was later validated and enriched by studies using amplified fragment-length polymorphism (AFLP), marker gene phylogeny, multi-locus sequence analysis (MLSA), comparative genome hybridization (CGH) and core-genome phylogeny (Portier et al., 2006; Costechareyre et al., 2010; Lassalle et al., 2011, 2017; Mousavi et al., 2015), as recently reviewed by Flores-Felix et al. (2020). At the time of writing this communication, the Agrobacterium genus contained 15 named species, of which 11 had validly published names. Moreover, nine out of 15 described genomic species within the Agrobacterium biovar 1 are not yet formally named: G3 (this study), G5, G6, G7 (“Agrobacterium deltaense”), G8 (“Agrobacterium fabrum”), G13, G15 (“Agrobacterium viscosum”), G19 and G20 (Mafakheri et al., 2019).

Several hundred microbial strains were isolated and identified in an ongoing investigation to map microbial diversity of the International Space Station (ISS; Checinska Sielaff et al., 2019). The ISS is a hermetically sealed closed system and their modules are at least 20-year-old, depending on their time of addition. The source of microorganisms is mainly through the human traffic, cargo transport and also associated with other experimental components. The air of the ISS is recirculated via an advanced environmental control system, and surfaces are maintained by implementing periodic cleaning. Previous attempts to isolate microorganisms from ISS environmental surfaces revealed that they generally consist of predominantly benign and commensal microorganisms (Mhatre et al., 2020; Bijlani et al., 2021), but potentially pathogenic microorganisms were also sporadically isolated (Knox et al., 2016; Singh et al., 2018a). In 2015, several bacterial strains were collected from the observation deck (Cupola) of the ISS, and whole-genome sequences (WGS) were generated for these isolates (Bharadwaj et al., 2020a,b; Bijlani et al., 2020a,b; Daudu et al., 2020a,b; Solomon et al., 2020; Simpson et al., 2021). Based on preliminary genomic analyses, three ISS strains were identified as belonging to the Agrobacterium genomospecies 3 (also known as A. tumefaciens genomovar G3 or Agrobacterium G3).

The first objective of this study is to describe the phylogenomic novelty and characterize the taxonomic affiliation of the three strains isolated from the ISS environment. We, therefore, assembled a dataset of complete genomes of Agrobacterium G3, including previously released sequences from five strains isolated from various geographical regions and environments: an eosin flask from a hospital environment in France (CFBP 6623, sequenced twice), a cave in Lechuguilla, United States (LC34), algae in the United Kingdom (SUL3), Arabidopsis roots in Germany (Root651), and bioprospecting for lignocellulolytic microbes and enzymes from natural, highly evolved plant biomass-degrading systems in the United States (UGM030330-04). We added two novel sequences from Agrobacterium G3 strains from our own collection, including an opportunistic pathogen isolated from cerebrospinal fluid (CFBP 6624) and one isolated from a tobacco plant rhizosphere (RTP8). In addition to traditional phenotype testing, molecular taxonomy utilizing 16S rRNA gene, gyrB, MLSA (gyrB, parE, recA, and rpoB genes), and

Keywords: Agrobacterium tomkonis, genomovar G3, ISS, metagenomics, phylogenomics, MLSA
core-gene-based phylogenetic analyses were carried out to describe Agrobacterium G3 strains.

Second objective of this study is to perform a comparative genomic analysis of Agrobacterium G3 strains together with representatives of other Agrobacterium species \((n = 41\) genomes) to elucidate genomic complexity. Subsequently, experimental characterization of the plasmid content was carried out on the set of Agrobacterium strains available to us in culture, i.e., G3 strains IIIF1SW-B1\(^T\), IIIF1SW-B3, IIIF1SW-B4, CFBP 6624, and RTP8 and reference strains "A. fabrum C58", Agrobacterium radiobacter B6\(^T\), and Agrobacterium pusense CFBP 5494. Thirdly, a pangenome analysis was implemented to identify core homologous gene clusters specific to ISS isolates and other Agrobacterium G3 strains. Finally, the metagenome reads of ISS environmental surfaces (Singh et al., 2018b) were mined for the presence of Agrobacterium G3 species to understand the prevalence of these novel species in ISS.

RESULTS AND DISCUSSION

The three strains collected from ISS belonging to Agrobacterium G3 along with six other G3 strains were subjected to a polyphasic characterization to determine the variable, conserved or distinctive traits of this genomic species.

Genome Characteristics of Novel Genome Sequences

Table 1 summarizes assembly statistics of all five novel ISS strains sequenced in this study. As an example, for strain IIIF1SW-B1\(^T\), the Illumina NovaSeq platform yielded \(1.8 \times 10^7\) paired-end (PE) reads were reduced to \(1.79 \times 10^7\) PE reads after performing trimming and quality filtering. The assembled draft genome of ISS strains consisted of 76–81 contigs with a genome size of \(6.25–6.29 \times 10^6\) bp; the assembled genomes of CFBP 6624 and RTP8 had 35 and 38 contigs, for a size of \(5.48 \times 10^6\) bp and \(5.44 \times 10^6\) bp, respectively. The contig N50 size was \(3.21–3.78 \times 10^3\)-bp, with a mean coverage of 100x for ISS strains, and 18x and 19x for strains CFBP 6624 and RTP8, respectively. G + C% of the five genomes was 59.12–59.21.

<table>
<thead>
<tr>
<th>Genome Characteristics</th>
<th>IIIF1SW-B1(^T)</th>
<th>IIIF1SW-B3</th>
<th>IIIF1SW-B4</th>
<th>CFBP 6624</th>
<th>RTP8</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCBI WGS accession</td>
<td>JABXYF0000000000</td>
<td>JABXYG0000000000</td>
<td>JABXYG0000000000</td>
<td>JARFL0000000000</td>
<td>JAFIRM0000000000</td>
</tr>
<tr>
<td># contigs (( \geq 0) bp)</td>
<td>137</td>
<td>154</td>
<td>150</td>
<td>44</td>
<td>55</td>
</tr>
<tr>
<td># contigs ((\geq 1000) bp)</td>
<td>76</td>
<td>81</td>
<td>80</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>Total length ((\geq 1000) bp)</td>
<td>6,234,019</td>
<td>6,273,081</td>
<td>6,273,787</td>
<td>5,480,209</td>
<td>5,440,309</td>
</tr>
<tr>
<td>Largest contig</td>
<td>644,466</td>
<td>644,466</td>
<td>726,576</td>
<td>1,342,491</td>
<td>928,660</td>
</tr>
<tr>
<td>Total length</td>
<td>6,240,486</td>
<td>6,282,603</td>
<td>6,282,746</td>
<td>5,483,735</td>
<td>5,445,056</td>
</tr>
<tr>
<td>GC (%)</td>
<td>59.18</td>
<td>59.17</td>
<td>59.17</td>
<td>59.12</td>
<td>59.21</td>
</tr>
<tr>
<td>N50</td>
<td>350,004</td>
<td>350,004</td>
<td>354,854</td>
<td>378,069</td>
<td>320,860</td>
</tr>
<tr>
<td># of genes</td>
<td>5,917</td>
<td>5,962</td>
<td>5,964</td>
<td>5,223</td>
<td>5,147</td>
</tr>
<tr>
<td># CDS</td>
<td>5,863</td>
<td>5,908</td>
<td>5,910</td>
<td>5,175</td>
<td>5,098</td>
</tr>
</tbody>
</table>

Phylogenetic Relationship of G3 Strains With Other Agrobacterium Species

A Maximum Likelihood (ML) tree of 16S ribosomal RNA (rRNA) gene sequences shows that this marker is not able to resolve and distinguish all Agrobacterium species (Figure 1A). All strains from genomospecies G3 and G4; G7 and G13; as well as G2 and G9 were found to have identical 16S rRNA gene sequences. In addition, the support for the topology of the 16S tree is generally low, with only branches leading to species A. rubi, Agrobacterium bohemicum and the clade formed by Agrobacterium genomospecies G6 and G8 being well supported. In addition, the type strains of non-biovar 1 species A. larrymoorei and "Agrobacterium albertimagini" are positioned within the clade containing otherwise only Agrobacterium biovar 1 strains, indicating that these 16S alleles are likely the result of horizontal gene transfer events. Similar results showing high similarities or identical 16S rDNA sequences for Agrobacterium roseae NCIPB-1650\(^T\), Agrobacterium skierniewiczenie Di1472 (99.7%), and A. rubi NBRC 13261\(^T\) (99.5%) were reported (Kuzmanović et al., 2018). This altogether suggests that the 16S rRNA gene is not a good marker to identify Agrobacterium species or to study their relationships, as observed in other taxa (La Duc et al., 2004).

The gene gyrB provides a much better phylogenetic resolution of relationship among Agrobacterium species, with all except genomospecies G7 forming highly supported clades (Figure 1B). All nine strains belonging to G3 were grouped into one clade. The shallow grouping of species into clades is highly supported except G4, G7, and G9 strains, but deeper relationships within the Agrobacterium biovar 1 are not. This indicate that gyrB is a bona fide marker gene to identify and study the diversity of Agrobacterium species, even though its use as a marker for amplicon-based surveys has been shown to be impractical due to the presence of the paralogous gene parE in the genome rendering selective amplification difficult (Pulawska and Kahužna, 2012).

The MLSA tree provides further resolution for Agrobacterium species, where strains belonging to various genomospecies were placed in a tight clade with high bootstrap values (Figure 1C). The relationships between species according to this MLSA tree are mostly in agreement with previous reports based on the
FIGURE 1 | Phylogeny of 40 Agrobacterium spp. strains based on 16S rRNA gene, gyrB, and MLSA. Conserved marker gene phylogenies of 40 Agrobacterium distinct strains (41 when including the two published genome versions for strain CFBP 6623. (A) 16S rRNA gene phylogeny, based on 1,602 aligned positions), rootied with sequence from Rhizobium leguminosarum USDA 2370T; (B) gyrB gene phylogeny, based on 2,477 aligned positions; (C) multi-locus sequence analysis phylogeny, based on the concatenated alignments of genes parE, gyrB, recA, and rpoB, resulting into 9,968 aligned positions. All gene sequences were extracted from the 41 studied and then aligned with Clustal Omega (for coding genes, alignment was performed at the protein level and then reverse-translated into codons). Maximum-likelihood trees were inferred with RAxML-NG 1.0.0 under the model GTR + FO + G4m, taking the best of 20 independent inferences, started with 10 random trees and 10 parsimony-optimized tree. Branch supports were estimated with 200 Felsenstein bootstrap trees. Only branch support over 70% are displayed; full information on the trees are available at https://doi.org/10.6084/m9.figshare.14792148; https://doi.org/10.6084/m9.figshare.14792169; https://doi.org/10.6084/m9.figshare.14792100.

concatenation of multiple loci or core-genome genes (Mousavi et al., 2014, 2015; Lassalle et al., 2017), indicating that this four-loci MLSA scheme (including gyrB, parE, recA, and rpoB genes) provides an efficient way to affiliate phylogenetic relationships among Agrobacterium species. All nine strains of Agrobacterium G3 formed a clade with a bootstrap value > 88%.

Finally, a tree based on the concatenation of 1,089 core-genome genes is fully resolved, with almost all branches having the highest support (Figure 2, bootstrap supports indicated if not 100%), and thus constitutes the gold standard for depicting the phylogenetic relationships of Agrobacterium species.

According to the gyrB and MLSA trees, the closest species to Agrobacterium G3 strains is the group formed of "A. fabrum" (G8) and Agrobacterium G6 strains, and these results were in accordance with previous reports (Lassalle et al., 2017). However, the core-genome tree revealed that the closest relatives were strains Agrobacterium sp. (Rhizobium sp.) UGM0330-04 and Agrobacterium sp. (Agrobacterium rhizogenes) K599 whereas strains belonging to genomovars G8 and G6 formed a different and distant clade (Figure 2). We note that strains UGM030330-04 and K599 are misnamed given these phylogenies clearly show that they are members of the Agrobacterium genus and thus not...
part of *Rhizobium*, nor of *A. rhizogenes*, which is a synonym of *R. rhizogenes*; they are thus more correctly referred to as *Agrobacterium* sp.

### Overall Genome Relatedness Indexes

All average nucleotide index (ANI) values among *Agrobacterium* G3 strains were over 97.5% whereas when comparing G3 strains to non-G3 strains, the ANI values were below 90.68% (Table 2 and Supplementary Table 1). The ANI values of *Agrobacterium* G3 strains and most closely related (as per core-genome) *Agrobacterium* sp. strains UGM030330-04 and K599 were ~90.5% similarity. Similarly, the ANI values of *Agrobacterium* G6 strains with other closely related – according to *gyrB* and MLSA trees – *Agrobacterium* G6 (NCPPB 925) and G8 strains (J-07 and C58), exhibited ~87.6% relatedness. Furthermore, average amino-acid identity (AAI) values among...

![Phylogeny of 40 Agrobacterium spp. strains based on the concatenation of 1,089 core-genome genes.](image-url)
Singh et al. Comparative Genomics of Agrobacterium tomkonis

<table>
<thead>
<tr>
<th>Organism</th>
<th>GenBank #</th>
<th>OGRI values in comparison to <em>A. tomkonis</em> IIF1SW-B1T</th>
<th>Genomospecies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ANI</td>
<td>AAI</td>
</tr>
<tr>
<td><em>Agrobacterium</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Agrobacterium</em> tomonis IIF1SW-B1T</td>
<td>JABXYF000000000</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td><em>Agrobacterium</em> tomonis IIF1SW-B3</td>
<td>JABXYG000000000</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td><em>Agrobacterium</em> tomonis IIF1SW-B4</td>
<td>JABXYH000000000</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 CFBP 6623</td>
<td>GCF_900013535.1</td>
<td>98.08</td>
<td>97.62</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. (Rhizobium sp.) Root651</td>
<td>GCF_001427625.1</td>
<td>97.74</td>
<td>97.25</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 LC34</td>
<td>GCF_001009581.1</td>
<td>97.82</td>
<td>97.77</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 SUL3</td>
<td>GCF_001263296.1</td>
<td>97.64</td>
<td>97.73</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 CFBP 6624</td>
<td>JAFIRL000000000</td>
<td>98.00</td>
<td>98.09</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 RTPB</td>
<td>JAFIRM000000000</td>
<td>98.08</td>
<td>98.15</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 CFBP 6623</td>
<td>GCA_005221385.1</td>
<td>98.12</td>
<td>97.61</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. (Rhizobium sp.) UGM030330-04</td>
<td>GCF_001427625.1</td>
<td>90.58</td>
<td>93.91</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. (<em>A. rhizogenes</em>) Root651</td>
<td>GCF_002005205.3</td>
<td>90.51</td>
<td>93.39</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 7 Zutra 3/1</td>
<td>GCF_900013515.1</td>
<td>98.63</td>
<td>91.71</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 LC34</td>
<td>GCF_001009581.1</td>
<td>97.82</td>
<td>97.77</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 SUL3</td>
<td>GCF_001263296.1</td>
<td>97.64</td>
<td>97.73</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 CFBP 6624</td>
<td>JAFIRL000000000</td>
<td>98.00</td>
<td>98.09</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 RTPB</td>
<td>JAFIRM000000000</td>
<td>98.08</td>
<td>98.15</td>
</tr>
<tr>
<td><em>Agrobacterium</em> genomspecies 3 CFBP 6623</td>
<td>GCA_005221385.1</td>
<td>98.12</td>
<td>97.61</td>
</tr>
</tbody>
</table>

*Agrobacterium* G3 strains were over 97.6% while values between G3 and other agrobacteria were below 94%. Likewise, the digital DNA:DNA hybridization (dDDH) values among *Agrobacterium* G3 strains were over 83.6% while values between *Agrobacterium* G3 and other agrobacteria were 41.7% or below. More specifically the dDDH values were ~41.7% for comparisons with strains UGM030330-04 and K599, 34.5% with G6 strains, and <33.9% with G8 strains. The ANI and dDDH values obtained for all the *Agrobacterium* G3 strains – including the ISS isolates – with other *Agrobacterium* species were below the threshold of 95% ANI (Yoon et al., 2017) and 70% dDDH values (Auch et al., 2010), which were established as standard for prokaryotic species delineation. We thus propose the name *Agrobacterium* tomonis for the genomspecies 3 (= genomovar G3) of the *Agrobacterium* biovar 1 (i.e., *A. tumefaciens* species complex).

*Agrobacterium* sp. strains UGM030330-04 and K599 consistently form a distinct group in the gyrB, MLSA and
core-genome tree; these two strains share an ANI of 98.11% and have <90.68% ANI with any other Agrobacterium strain. We therefore propose to group them into a new genomospecies of the Agrobacterium biovar 1, to be named genomovar G21, following the recently described genomovars G19 and G20 (Mafakheri et al., 2019).

Phenotypic Characterization
Differential phenotypic characteristics of A. tomkonis strains with six other Agrobacterium species are given in Table 3 and Supplementary Table 2. The biochemical tests were highly variable among Agrobacterium species tested, in accordance with previous reports (Kuzmanovi´c et al., 2018). However, we found a distinctive trait in that all Agrobacterium biovar 1 strains except those belonging to the A. tomkonis assimilated D-galacturonic acid. In addition, A. tomkonis strains tested assimilated the following carbon sources (Supplementary Table 2). D-maltose, D-trehalose, D-cellulose, gentiobiose, sucrose, D-turanose, stachyose, D-raffinose, α-D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, N-acetyl-D-galactosamine, α-D-maltose, D-malic acid, bromosuccinic acid, L-lactic acid, D-malic acid,手套状酸, D-Galacturonic Acid, N-Acetyl-D-Galactosamine, Methyl Pyruvate, p-hydroxyphenylacetic acid, and sodium butyrate.

Among various compounds tested, ISS A. tomkonis strains did not metabolize citric acid and Niaproof 4 (an anionic surfactant) whereas other A. tomkonis strains utilized them. Most of the compounds tested were metabolized unvaryingly by all ISS strains whereas strain IIF1SW-B3 was able to utilize lithium chloride as sole carbon source as that of other A. tomkonis strains tested. Similarly, strain IIF1SW-B3 was metabolizing TWEEN 40 (polysorbate, surfactant) but other A. tomkonis strains did not assimilate it. However, all six tested A. tomkonis strains are tolerant to troleandomycin, rifamycin SV, lincomycin, guanidine HCl, vancomycin, potassium tellurite, aztreonam, 1% sodium lactate, tetracil violet, and tetracil blue, but not to minocycline. Moreover, they all grew at pH 6 but not or only weakly at pH 5. All six tested A. tomkonis strains did not grow at 4% NaCl and beyond. The strains IIF1SW-B1, IIF1SW-B4, and RTP8 tested on tomato produced no tumor (results not shown) and are therefore considered not pathogenic on tomato plant.

### Comparative Genomic Analysis and A. tomkonis-Specific Gene Content
Forty-nine genes were specifically found in A. tomkonis genomes. These genes were mostly interspersed in the genome and coded a range of functions (Supplementary Table 3), but a fraction clustered into three genomic islands, two on the linear chromosome and one on the largest of the two megaplasmids present in strain CFBP 6623 (pCFBP6623a). The first cluster of seven A. tomkonis specific genes clearly encodes the biosynthesis of a surface polysaccharide; the second encodes mostly proteins of unknown function, with one potentially related to type 6 secretion systems; the last A. tomkonis specific gene cluster on the plasmid encodes resistance to copper. Other A. tomkonis specific genes are scattered in the genome, but have related functions. There are notably two genes encoding ankyrin-repeat domain-containing proteins, as well as a protein containing an extracellular HAF repeat domain (InterPro domain IPR014262).

### Table 3 | Differential biochemical characteristics between A. tomkonis and closely related species.

<table>
<thead>
<tr>
<th>Biochemical tests</th>
<th>A. tomkonis IIF1SW-B1</th>
<th>A. tumefaciens CFBP 5771</th>
<th>A. neoptum 39/7</th>
<th>“A. fabrum” C58</th>
<th>A. pusense CFBP 5494</th>
<th>A. radiobacter B6</th>
<th>A. larrymoorei AF3.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Raffinose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>α-D-Lactose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>L-Arginine</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>Methyl Pyruvate</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>w</td>
<td>w</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N-Acetyl-D-Galactosamine</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>w</td>
<td>w</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>D-Glucuronic Acid</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>D-Saccharic Acid</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>D-Glucuronic Acid</td>
<td>w</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>L-Lactic Acid</td>
<td>w</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>D-Malic Acid</td>
<td>w</td>
<td>+</td>
<td>−</td>
<td>w</td>
<td>+</td>
<td>+</td>
<td>nd</td>
</tr>
<tr>
<td>Bromosuccinic Acid</td>
<td>w</td>
<td>w</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Acetoacetic Acid</td>
<td>w</td>
<td>w</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>nd</td>
</tr>
<tr>
<td>Pectin</td>
<td>w</td>
<td>+</td>
<td>−</td>
<td>w</td>
<td>−</td>
<td>−</td>
<td>nd</td>
</tr>
</tbody>
</table>

*Data from Puławska et al. (2012); **Data from Bouzar and Jones (2001), Pandey et al. (2011). +, positive; −, negative; w, weak reactions; nd, not determined.
and an autotransporter domain, all three exported to the cell surface and likely mediating attachment to an extracellular substrate. Together with the biosynthesis of a likely extracellular polysaccharide, these species-specific genes present a convergent function related to surface attachment and potentially resistance to stress from the external medium. This enrichment in the specific core-genome of *A. tomponis* suggests natural selection was at work and that surface attachment may have a strong role in *A. tomponis* ecology. Indeed, the first reported *A. tomponis* strains CFBP 6623 and CFBP 6624 were isolated from within an eosin flask and from a human clinical sample (likely as an opportunist taking advantage of another pathogen infection), respectively (Popoff et al., 1984), showing already unusual colonization abilities for agrobacteria. Moreover, the strains recovered from the ISS were isolated from the surface of the observation dome (Cupola), which panel is made of aluminum with a polyurethane topcoat (e.g., Aeroglaze A276 or BMS10-60), an ultra-clean and smooth surface that carries very little microbial life. The independent isolation of G3 strains in such locations suggest that attachment to inhospitable surfaces may be a way for *A. tomponis* to colonize new habitats where it can evade competition with other microbes.

Similarly, the gene clustering analysis between the genomes of 40 different *Agrobacterium* strains allowed the identification of ISS strains-specific gene clusters (Figure 3 and Supplementary Table 4). We found 556 genes specifically present in the three ISS strains; we identified these genes based on their presence in all ISS strains and none of the 17 genomes in the background clade comprising the other *A. tomponis* strains, all strains of *Agrobacterium arsenijevicii*, *Agrobacterium nepotum*, *Agrobacterium genomospecies* G1, G5, G13, and G2 (Supplementary Data online). Using a more relaxed criterion allowing presence of the genes in a maximum of 2 out the 17 genomes in the background clade, to allow the detection of genes specifically gained by the ancestor of the ISS strains but that may have independently introduced in close relatives via horizontal gene transfer, we found 838 genes (Supplementary Table 4), most of them located in large contiguous regions.

Among this wider set of ISS clone-specific genes, some were found to cover 10 entire contigs of the genome assembly of strain IIF1SW-B1T (Supplementary Table 5), totaling to 347 genes over 613 kb of ISS clone-specific contigs, which possibly represent (parts of) extrachromosomal elements. Indeed, four of these ISS clone-specific contigs carried genes with functions associated to type 4 secretion systems (T4SS), with two seemingly complete clusters, indicating they may constitute conjugative elements such as plasmids or integrative and conjugative elements (ICE). Most of the other ISS clone-specific contigs carried transposase or genetic mobility-associated genes, suggesting they also constitute mobile genetic elements (MGEs). In addition, large contiguous portions of other contigs that were specific to the ISS strains carried transposase genes – often located at the extremity of the ISS clone-specific region – suggesting they may represent inserted MGEs such as transposons or genomic islands (Supplementary Tables 4, 5). In particular, the ISS clone-specific region of contigs #27 of strain IIF1SW-B1T genome assembly clearly corresponds to a prophage.

Beyond genetic mobility, the ISS clone-specific genes displayed a range of other functions, which were enriched for a range of functions, some forming coherent cellular processes (Supplementary Table 6). Among the top enriched functions, we found a multi-copper oxidase, accompanied by a chaperone for its maturation and transporters that might allow the assimilation of copper ions from the environment; these functions may form a pathway for biosynthesis of a redox enzyme that could be involved in a respiratory chain. Another set of coherent functions were found to be involved in scavenging and uptake of a ferric ion-siderophore complex, as well as a non-ribosomal peptide synthases-polyketide synthases (NRPS/PKS) biosynthetic cluster, which might be involved in the biosynthesis of the siderophore.

**Plasmid Profiling**

Based on migration of genomic DNA on Eckhardt gels (Vaudequin-Dransart et al., 1995), all *A. tomponis* strains were evidenced to carry plasmids, although differences in size and number were observed between strains (Supplementary Figure 1). Indeed, strain CFBP 6624 presented a single plasmid of approximately 500 kb, whereas CFBP 6623, RTP8 and the three ISS strains showed one to three plasmids, with a common largest band around 260 kb. RTP8 and the three ISS strains have in common a 130-kb plasmid, and a 40-kb plasmid is present exclusively in the three ISS strains. The ISS strains do not harbor a pTi plasmid, as no classical *vir* genes were detected in their genomes (data not shown). Accordingly, no symptoms of crown-gall diseases were observed on injured tomato plants inoculated with any of the ISS strains (data not shown).

**Agrobacterium tomponis** Reads in International Space Station Metagenomes

In order to evaluate the relative abundance of *A. tomponis* onboard the ISS, we performed metagenomic read recruitment obtained from ISS environmental DNA shotgun datasets (Singh et al., 2018b) with strain IIF1SW-B1T genome as a reference. Visual inspection of the mapped read profiles showed that most recruited reads were located on conserved bacterial genes encoding products such as ribosomal RNA operons, ribosomal proteins, DNA-directed RNA polymerase *rpoB* or elongation factor Tu. Considering the high degree of conservation of these genes across distant bacterial lineages and the non-uniform read coverage of G3 contigs we conclude that despite being isolated from the same environment, *A. tomponis* is not present in any detectable abundance in the ISS metagenomes.

In summary, *A. tomponis* genomes are diverse, and notably carry a variety of plasmids, even though none of the G3 strain in our dataset carried a tumorigenic plasmid. As previously shown for other *Agrobacterium* genomic species (Lassalle et al., 2017), *A. tomponis* strain genomes share a number of species-specific genes. *A. tomponis* specific gene functions notably relate to surface adhesion and could be involved in the ecological specificity of *A. tomponis* suggested by their varied source of

---

1https://10.6084/m9.figshare.16782964
isolation (Popoff et al., 1984). Indeed, while most species of the *Agrobacterium* biovar 1 are usually isolated from soils, plant rhizospheres or as plant pathogens inducing tumors (Barton et al., 2018; Dodueva et al., 2020; Weisberg et al., 2020), *A. tomkonis* strains never have been isolated as a plant pathogen, nor a tumor-inducing plasmid was observed in their genome, which we confirmed by negative infection tests on tomato plants. Even though the isolation of *A. tomkonis* strains RTP8 from a tobacco plant rhizosphere and Root651 from *Arabidopsis* roots prove that *A. tomkonis* strains are able to survive in plant rhizospheres, *A. tomkonis* strains were otherwise all isolated from a variety of un hospitable environments, including from an antiseptic flask, as an opportunistic pathogen of the human central nervous system, from a cave wall, and lastly from the inert surface of the ISS cupola. The ability to colonize these substrates, unusual for most agrobacteria, resonate with the finding that *A. tomkonis* genomes specifically carry genes involved in mediating attachment to surfaces, including production of putative adhesins and biofilm. This suggests a particular ability of *A. tomkonis* to colonize a different kind of habitats, that are even...
poorer in nutrients and harsher than soil, thus possibly escaping competition with other agrobacteria that are better at growing in richer environments such as plant rhizospheres.

The genomes of *A. tomkonis* strains isolated from the ISS, appear to be highly similar, suggesting they are clonally related. As such, they carry further specific genomic traits, including the presence of a 40-kbp plasmid observed via electrophoresis, as well as several other potential MGEs detected in their specific genomic content, with several ISS clone-specific contigs carrying conjugation genes (cumulated length 346 kbp) and more contigs that could also be part of conjugative elements or other types of MGEs such as integrated prophages. These ISS clone-specific genes also include many potentially adaptive genes, including putative pathways for respiratory chain biosynthesis and iron scavenging, but it is unclear what would be the selective advantage those genes could bring to the ISS strains in the context they were isolated from.

**CONCLUSION**

Using a polyphasic approach, we characterized the phenotypic and genotypic synapomorphies of *Agrobacterium* G3 (*Agrobacterium* genospecies 3), showing it is a *bona fide* bacterial species, well separated from previously named *Agrobacterium* species or other recognized genomic species. Based on this evidence, we propose to name this species *Agrobacterium tomkonis* (formal description below). In addition, we performed an in-depth investigation of the pangenome, notably identifying species-specific genes, allowing us to relate the description of this new species to hypotheses on its ecology. Within this novel species, we further characterized the clonal group of strain isolated from the ISS, showing it possesses its own set of specific genes and MGEs that could carry adaptive traits linked to its survival in the ISS environment.

**Description of Agrobacterium tomkonis** sp. nov.

*Agrobacterium* *tomkonis* (tom. ko’ni.s. N.L gen. n. *tomkonis* referring to David Tomko, a well-known NASA Space Biology scientist who advanced space research in the United States). Cells are Gram-stain-negative, non-spore-forming, aerobic rods, 0.5–0.6 μm in width and 0.8–1.6 μm in length. Colonies are translucent and cream to white, with a diameter of 2–3 mm on R2A medium after incubation for 3 days at 25°C. Growth occurs at 20°C through 35°C, with an optimum temperature of 25°C. *A. tomkonis* strains grew at pH 6 but not or only weakly at pH 5 with a pH optimum at 7.5. All tested *A. tomkonis* strains failed to grow at 4% NaCl and beyond.

*A. tomkonis* strains metabolize D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, D-turanose, stachyose, D-raffinose, α-D-lactose, melibiose, β-methyl-D-glucoside, D-salicyl, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, α-D-glucose, D-mannose, D-fructose, D-galactose, D-fucose, L-fucose, L-rhamnose, D-sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol, D-glucose- 6-PO4, D-fructose- 6-PO4, L-alanine, L-arginine, L-aspartic acid, L-histidine, L-lysoglutamic acid, L-serine, pectin, D-gluconic acid, D-Glucuronic acid, quinic acid, methyl pyruvate, L-malic acid, γ-amino-butyric acid, propionic acid, acetic acid. However, did not utilize N-acetyl-β-D-mannosamine, N-acetyl neuraminic acid, 3-Methyl Glucose, fusidic acid, D-serine, D-aspartic acid, D-galacturonic acid, L-galactonic acid lactone, mucic acid, D-saccharic acid, p-hydroxyphenylactic acid, and sodium butyrate.

*A. tomkonis* strains tested are tolerant to troleandomycin, rifamycin SV, lincomycin, guanidine HCl, vancomycin, potassium tellurite, aztreonam, 1% sodium lactate, tetrazolium violet, and tetrazolium blue, but not to minocycline.

The delineation of *A. tomkonis* from other *Agrobacterium* species was based on overall genome relatedness indexes (all *A. tomkonis* share ANI > 98% and dDDH > 84%), as well as based on well-supported clades in gyrB gene, MLSA (with gyrB, parE, recA, and rpoB genes) or core-genome phylogenies. The type strain, IIF1SW-B1T (= LMG 32164 = NRRL B-65602), was isolated from the ISS Port panel of the Cupola, which is the observation deck for the crew. The type strain DNA G + C content is 59.18 mol% and its genome sequence is available from GenBank under WGS accession JABXYF000000000.

**MATERIALS AND METHODS**

**Sample Collection and Isolation of Bacteria**

The sampling of ISS surfaces performed for this study took place within the United States on-orbit segments. Samples collected during this study were: Node 3 (Locations #1, #2, and #3), Node 1 (Locations #4 and #5), Permanent Multipurpose Module (Location #6), U.S. Laboratory (Location #7), and Node 2 (Locations #8 and control). A detailed description of the various locations sampled was published elsewhere (Singh et al., 2018b). Sample collection from ISS environmental surfaces, processing, and cultivation of bacteria have been previously reported (Checinska Sielaff et al., 2019). Three strains isolated during the second flight from Location #1 (Cupola) surfaces, were later identified as belonging to the genus *Agrobacterium* via 16S rRNA gene sequencing of the cultivated strains; these were designated: IIF1SW-B1T, IIF1SW-B3, and IIF1SW-B4.

**Phenotypic Characterization**

Strains IIF1SW-B1T, IIF1SW-B3, IIF1SW-B4, CFBP 6623, CFBP 6624, and RTP8 strains were subjected for the phenotypic characterization. All strains are grown on yeast extract-peptone-glucose (YPG) medium and collected from the surface of the agar plates (Scortichini, 2012) before inoculation and screened for various organic substrate utilization using Biolog GEN III system (Biolog, Inc., Hayward, CA, United States) following manufacturer’s recommendations. Concurrently, two replicates of Biom tests were carried out, and Biom plates were incubated at 28°C for 48 h. After incubation, the OmniLog™ system was used to measure carbon substrate utilization.
Molecular Characterization of International Space Station Strains

A loopful of purified microbial culture (only ISS strains) was subjected to DNA extraction with the UltraClean DNA kit (MO BIO, Carlsbad, CA, United States) or Maxwell Automated System (Promega, Madison, WI, United States) as per manufacturer instructions. The extracted DNA was eluted in 50-µL molecular grade water and stored at −20°C until further analysis. The 16S rRNA gene was amplified using the forward primer, 798F (5′-AGA GTT TGA TCC TGG CTC AG-3′) and the reverse primer, 1492R (5′-GTT TAC CTT GTT ACG ACT T-3′) (Checinska Sielaff et al., 2019) and PCR was performed with the following conditions: denaturation at 95°C for 5 min, followed by 35 cycles consisting of denaturation at 95°C for 50 s, annealing at 55°C for 50 s, and extension at 72°C for 1.5 min, and finalized by extension at 72°C for 10 min. The amplified products were treated with Antarctic phosphatase and exonuclease (New England Biolabs, Ipswich, MA, United States) to remove 5′- and 3′- phosphates from unused dNTPs before sequencing. Sequencing was performed by Macrogen (Rockville, MD, United States) using 27F and 1492R primers for Bacteria. The resulting sequences were assembled using SeqMan Pro from the DNASTAR Lasergene package (DNASTAR Inc., Madison, WI, United States). Bacterial sequences were searched against the EzTaxon-e database (Kim et al., 2012) and identified based on the closest percentage similarity (>98.6%) to previously identified microbial-type strains.

Analysis of Plasmid Content

Plasmid profiles were determined by a modified Eckhardt agarose gel electrophoresis technique as previously described (Vial et al., 2006). Isolates were grown overnight in YPG medium at 28°C until the optical density at 600 nm reached 0.4. Electrophoresis was carried out at 5 V for 30 min and 90 V for 5 h at 4°C on a 0.75% agarose gel containing 1% sodium dodecyl sulfate. Plasmids sizes were estimated by comparison with those of “A. fabrum” C58 and Allorhizobium vitis S4 (Wood et al., 2001; Slater et al., 2009).

Analysis of International Space Station Strains Pathogenicity on Tomato Plant

To test if the ISS G3 strains were plant-pathogen strains, 3-week-old tomato plants, cultivated with a photoperiod of 16 h/8 h light/dark in a greenhouse were stem-injured and inoculated with 10 µL of bacterial overnight culture (10⁸ CFU/ml). The plants were then incubated for 21 days, and the presence or absence of crown gall symptoms was established by visual inspection. “A. fabrum” C58 strain was used as the positive control.

Whole Genome Sequencing of New Agrobacterium Strains

The WGS sequencing of the three ISS G3 strains was carried out as per established procedures (Singh et al., 2018a). Shotgun libraries were prepared using the Illumina Nextera Flex protocol (Singh et al., 2018b) and sequenced using NovaSeq 6000 S4 flow cell 2 × 150 PE sequencing kit. Verification of the quality of the raw sequencing data was carried out using FastQC v0.11.7 using default parameters². Quality control for adapter trimming and quality filtering were performed using fastp v0.20.0 (Chen et al., 2018), and then SPAdes v3.11.1 (Bankevich et al., 2012) was used to assemble all the cleaned sequences. Fastp quality control was based on the following three parameters: (i) correction of mismatches in overlapped regions of paired-end reads, (ii) trimming of autodetected adapter sequences, and (iii) quality trimming at the 59 and 39 ends. To determine the quality of the assembled sequences, the number of contigs, the N50 value, and the total length were calculated using QUAST v5.0.2 (Gurevich et al., 2013). Default parameters were used for all software. The WGS sequencing of strains RTP8 and CFBP 6624 was carried out by Beckman Coulter Genomics (Takeley, Essex, United Kingdom). SPAdes v3.8.1 (Bankevich et al., 2012) was used to assemble all the cleaned sequences. All five new genomes were annotated using Prokka v1.14.5 (Seemann, 2014) with the following options: “--force --addgenes --compliant --usegenus.” Protein function was annotated by performing a blastp similarity search against a reference proteome database made of 405 high-quality genome assemblies obtained from the NCBI RefSeq, selecting all genomes belonging to the Rhizobiales family available as on 14th August 2018 that had a contig N50 greater than 98 kbp (Supplementary Table 7).

Overall Genome Relatedness Index Calculations

The ANI (Yoon et al., 2017) between each pair of the 40 Agrobacterium genomes (41 genomes including the two published versions of strain CFBP 6623 genome) was calculated using pyANI (Pritchard et al., 2016). The dDDH analysis was performed using the Genome-to-Genome Distance Calculator 2.0 (GGDC 2.0) (Auch et al., 2010).

Pangenome Analysis

We applied two separate pangenome analysis pipelines to obtain robust estimates of the sets of genes present and absent in each genome and to compute core and clade-specific gene sets for the various taxonomic groups represented in our dataset, including Agrobacterium biovar 1, A. tomkonis strains including ISS isolates.

(1) On one hand, a pangenomic Snakemake workflow (Köster and Rahmann, 2012) was run in anvi’o v6.2 (Eren et al., 2015). Briefly, gene clusters (GC) were defined by clustering all-versus-all open-reading frames BLAST similarity scores using the Markov Clustering (MCL) algorithm (Enright et al., 2002) with an inflation score of 8. Open reading frames were identified with Prodigal 2.6.3 in single mode (Hyatt et al., 2010). Open reading frames were annotated using COG and KOam databases (Eren et al., 2015; Delmont and Eren, 2018). Results were displayed using anvi-display-pan function (Delmont and Eren, 2018). Functional enrichment in the ISS and G3 strain sets was obtained using anvi-compute-functional-enrichment function, as described in Shaiber et al. (2020), based on the COG annotation

²http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
of open reading frames. Only enrichment for which FDR adjusted p-value < 0.05 are reported in Supplementary Table 5.

(2) On the other hand, a phylogenomic database was built with the Pantagruel pipeline (Lassalle et al., 2019), using the “usingGeneRax” branch of the code version “dl79feca577b72389926a0138353fa657c1f3f368.” The following pipeline tasks were performed: init, 00-03 and 05, leading to the clustering of homologous protein families with MMseqs2 (Steinenger and Söding, 2017), the alignment of their reverse-translated coding sequences with Clustal Omega (Sievers et al., 2011) and Python scripts using Biopython (Weisberg et al., 2020), and the computation of a maximum-likelihood (ML) phylogenetic tree with RAxML 8.1.2 (Stamatakis et al., 2005) based on the concatenation of 1,089 core-genome gene alignments, as previously described (Lassalle et al., 2021) (Supplementary Data File doi: 10.6084/m9.figshare.14792178).

In addition, clade-specific sets of homologous gene clusters were determined using the “get_clade_specific_genes.r” script from Pantagruel package using the homologous gene family presence/absence matrix as input.

Metagenome Sequence Reads

Recruitment to Strain IIF1SW-B1

Metagenomic datasets (n = 42) previously obtained from environmental swabbing of the ISS (Singh et al., 2018b) were downloaded from NCBI-SRA BioProject PRJNA438545, quality filtered and mapped using BWA-MEM (Li and Durbin, 2009) against the genome assembly of strain IIF1SW-B1\textsuperscript{T}. These steps were performed using anvi’o v6.2 (Eren et al., 2015) using anvi’o metagenomic workflow in reference mode and default parameters. Visual inspection of read mapping alignment was carried out using Artemis BamView (Carver et al., 2012).

Phylogenetic Analysis of Marker Genes

For the 16S rRNA gene phylogenetic analysis, we extracted the 16S rRNA sequences from the genomes and added the reference 16S gene sequence from \textit{Rhizobium leguminosarum} type strain USDA 2370\textsuperscript{T} (GenBank accession U29386.1) and then aligned all sequences with Clustal Omega (Sievers et al., 2011) (Supplementary Data File doi: 10.6084/m9.figshare.14792148). For the \textit{gyrB} single-gene phylogenetic analyses, as well as multi-locus sequence analysis (MLSA), we used the coding sequence (CDS) alignments for the marker genes \textit{parE}, \textit{gyrB}, \textit{recA} and \textit{rpoB} from the Pantagruel database described above (see Supplementary Data Files available at doi: 10.6084/m9.figshare.14792169 and doi: 10.6084/m9.figshare.14792100). For the MLSA, we concatenated the CDS alignments of \textit{parE}, \textit{gyrB}, \textit{recA}, and \textit{rpoB} into a multi-locus alignment, totaling 9,968 positions. ML Phylogenies were then inferred based on these alignments with RAxML-NG 1.0.0 (Stamatakis et al., 2005) with the options “- model GTR + G4 -all,” implying the use of the model GTR + FO + G4m, and that 20 independent inferences were run, starting with 10 random trees and 10 parsimony-optimized tree, and retaining the best ML tree from these 20 inferences. Branch supports were estimated with 200 Felsenstein bootstrap trees. The 16S tree was rooted using \textit{R. leguminosarum} USDA 2370\textsuperscript{T} sequence, whereas the \textit{gyrB} and MLSA trees were rooted using the sequence from “A. albertimagni” AOL15 – a distant relative incorrectly classified in the \textit{Agrobacterium} genus for which a classification in the newly proposed genus “\textit{Petryoungia}” was proposed (Kuzmanović et al., 2021).

DATA AVAILABILITY STATEMENT

The draft genome sequences of three ISS strains including type strain IIF1SW-B1\textsuperscript{T}, as well as strains RTP8 and CFBP 6624, were deposited in NCBI. The genome sequences of isolates \textit{A. tomkonis} IIF1SW-B1\textsuperscript{T}, RTP8, and CFBP 6624 have been deposited to GenBank under the BioSample accessions SAMN15333190, SAMN17917486, and SAMN17918142, respectively; the WGS project accession numbers are also given in Table 1. The genome versions described in this paper are the first versions. Analysis intermediary data and result tables are available under the FigShare project available at https://figshare.com/account/home/#/projects/115866 with individual items available at the following DOIs: 10.6084/m9.figshare.14792100, 10.6084/m9.figshare.14792148, 10.6084/m9.figshare.14792169, 10.6084/m9.figshare.14792178, 10.6084/m9.figshare.14883546, 10.6084/m9.figshare.14883516, 10.6084/m9.figshare.14798289, and 10.6084/m9.figshare.16782964.

AUTHOR CONTRIBUTIONS

CL, LV, XN, KV, and NS conceived and designed the microbiological experiments. NS, CL, and LV performed the microbiological experiments. CL and LV carried out the phenotypic assays Biolog based biochemical characterization. CL performed the plant pathogenicity experiments. LV performed plasmid content analyses. XN managed the \textit{Agrobacterium} strain collection and the non-ISS strain genome sequencing project. NS and KV managed the ISS strain collection and the ISS strain genome sequencing project. FL, JN, and NS analyzed the data. NS analyzed the \textit{de novo} assemblies, including contig alignment, and annotation checks. FL performed phylogenetic and phylogenomic analyses. JN performed metagenomic analysis. FL and JN performed pangenome analyses. FL and KV wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with National Aeronautics and Space Administration. This research was funded by a 2012 Space Biology NNH12ZTT001N grant no. 19-12829-26 under
ACKNOWLEDGMENTS

We would like to thank Aleksandra Checinska-Sielaff for isolating the strain and Achintya Bharathwaj for initial phylogenetic analyses, astronomers Captain Terry Virts for collecting samples aboard the ISS and the Implementation Team at NASA Ames Research Center (Fathi Karouia) for coordinating this effort, and Ryan Kemp (Zymo Research Corp.) for extracting DNA and Dan Butler (Weill Cornell Medicine) for performing shotgun sequencing using the NovaSeq platform. We acknowledged the Jet Propulsion Laboratory supercomputing facility staff, notably Narendra J. Patel (Jimmy) and Edward Villaneuva, for their continuous support in providing the best possible infrastructure for BIGDATA analysis. © 2021 California Institute of Technology, Government sponsorship acknowledged. We would also like to thank Elise Lacroix for the plant cultures at the “Serre et Chambres Climatiques” platform of the FR BioEnViViS (UCBL) and L. Loiseau of PARMIC/CRB-EML platform of Ecologie Microbienne UMR CNRS 5557, UMR INRAE 1418.

REFERENCES


SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.765943/full#supplementary-material

Supplementary Figure 1 | Plasmid profiles of A. tomkonis strains. Strains IIF1SW-B1, IIF1SW-B3 and IIF1SW-B4 showed the same plasmidic profile. Plasmids sizes were estimated by comparison with those of A. fabrum CS8 and A. vitis S4 (data not shown). LC indicates the linear chromosome.

Supplementary Table 1 | ANI matrix of all tested Agrobacterium strains.

Supplementary Table 2 | Differential biochemical characteristics among A. tomkonis strains.

Supplementary Table 3 | COG function enriched in 9 Agrobacterium tomkonis (G3) strains, compared to the 31 other Agrobacterium genomes in the dataset.

Supplementary Table 4 | List of Agrobacterium clade-specific homologous gene families. Clade-specific genes are listed per clade; with each list starting by a hashtag and the clade id: “# cladeXX”. A. tumefaciens species complex/Agrobacterium biovar 1 is clade1; A. tomkonis is clade20; the clonal group of ISS strains is clade27. Genomes are listed using short codes, with the correspondence to strain names provided at the top.

Supplementary Table 5 | List of contigs specific to the clonal group of A. tomkonis strains isolated from the ISS and their characteristics. The count of genes with genetic mobility-related functions is based on the Prokka annotation of proteins.

Supplementary Table 6 | COG function enriched in 3 Agrobacterium tomkonis ISS strains, compared to the 37 other Agrobacterium genomes in the dataset.

Supplementary Table 7 | List of genome assemblies used as reference to annotate the new Agrobacterium genomes with Prokka.


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Singh, Lavire, Nesme, Vial, Nesme, Mason, Lassalle and Venkateswaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.