INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
18F–017/2018/20–5–1167–1174
DOI: 10.17957/IJAB/15.0638
http://www.fspublishers.org

Full Length Article

The Complete Genome Sequence of Cucumopine-Type *Agrobacterium rhizogenes* Strain K599 (NCPPB2659), A Nature's Genetic Engineer Inducing Hairy Roots

Xiaxiu Tong[1], Yafei Li[1], Taihe Xiang[1,2*], Zhehao Chen[1], Lilin Wang[1], Chengchao Zhang[1], Mi Feng[1] and Laiyang Wu[1]
[1]*College of Life and Environment Sciences, Hangzhou Normal University, Hangzhou 310036, China*
[2]*Zhejiang Provincial Key Laboratory for Genetic Improvement and Quality Control of Medicinal Plants, Hangzhou 310036, China*
[*]For correspondence: xthcn@163.com; xthcn@hznu.edu.cn

Abstract

Agrobacterium rhizogenes induces hairy roots from the infection site, when transferred DNA (T-DNA) in Ri plasmid inserted and integrated into host plant cell genome. Cucumopine-type *A. rhizogenes* strain K599 (NCPPB2659) has particularly high ability to induce hairy root. Here, using next-generation and third generation sequencing technologies, we *do novo* assembled the whole genome of K599. The genome is 5,114,385 bp and consists of one circular chromosome of 2,622,758 bp, one linear chromosome of 2,289,334 bp, and one Ri plasmid (pRi2659) of 202,292 bp. It contains 4716 protein-coding and 70 RNA genes, including 54 tRNAs, 12 rRNAs, and 4 ncRNAs. The genetic information of K599 exhibits the prediction of prophage regions and virulent genes, which is critical for DNA transfer to the host plant genome. © 2018 Friends Science Publishers

Keywords: *Agrobacterium rhizogenes*; K599; NCPPB2659; Genome sequence; Hairy root; Cucumopine
Abbreviations: BLAST: Basic local alignment search tool; *chv*: chromosomal virulence; COG: Clusters of orthologous group; CRISPR: Clustered regularly interspaced short palindromic repeats; CTAB: Cetyl-trimethylammonium bromide; *cus*: cucumopine synthesis; ncRNAs: Non-coding RNAs; NGS: Next-generation sequencing; Ri: Root inducing; *rol*: root loci; *orf*: open reading frame; T-DNA: Transferred DNA; Ti: Tumor inducing; TGS: Third generation sequencing; VFDB: Virulence factor database; *vir*: virulence; YEM: Yeast extract mannitol

Introduction

Agrobacterium is a Gram-negative soil bacterium and can cause pathogenic diseases in plants. There are two types of *Agrobacterium*: *A. tumefaciens* and *A. rhizogenes* (*Rhizobium rhizogenes*). During process of *Agrobacterium* infecting host plant, the transferred DNA (T-DNA) in Ti/Ri plasmid from *A. tumefaciens*/*rhizogenes* inserts and integrates into plant cell nuclei. Consequently, *A. tumefaciens* produces tumor while *A. rhizogenes* induces hairy root from the infection site (White and Nester, 1980; Chilton *et al*., 1982). Based on the infection process of T-DNA exported from *Agrobacterium* to the plant cell nucleus and expressed, *Agrobacterium*-mediated plant transformation was developed. At present, *Agrobacterium*-mediated transformation is the most widely used approach in plant genetic engineering and has revolutionized agriculture as well as basic research in plant molecular biology (Banta and Montenegro, 2008).

Agrobacterium strain has the ability to induce the synthesis of specific amino acid and sugar derivatives, called opines in host plant as its nutrients. According to opines produced in plants induced by its infection, *A. rhizogenes* is classified into four types (agropine, mannopine, cucumopine and mikimopine) (Combard and Baucher, 1988; Suzuki *et al*., 2001). K599, also known as NCPPB2659, is an archetype of a cucumopine-type *A. rhizogenes* strain. Infection by K599 induces hairy roots in 32 different plant species (Xiang *et al*., 2016), especially early recalcitrant leguminous plants and monocots like corn (Cho *et al*., 2000; Runo *et al*., 2012). In addition, K599 induces hairy roots in soybean with higher efficiency comparing to other strains (Savka *et al*., 1990; Cho *et al*., 2000), which can be harnessed for soybean genetic engineering.

Hairy root system induced by *A. rhizogenes* has been broadly utilized for basic and applied research (Guillon *et al*., 2006; Georgiev *et al*., 2012). The molecular mechanisms of processes such as T-DNA transfer, interaction with host plant proteins, plant defense signaling, and integration to the plant genome are partially known. For example, root loci genes (*rol* genes), virulence genes (*vir* genes) in Ri plasmid, and chromosomal virulence genes (*chv* genes) of *A. rhizogenes* are all involved in the hairy

To cite this paper: Tong, X., Y. Li, T. Xiang, Z. Chen, L. Wang, C. Zhang, M. Feng and L. Wu, 2018. The complete genome sequence of cucumopine-type *Agrobacterium rhizogenes* strain K599 (NCPPB2659), a nature's genetic engineer inducing hairy roots. *Int. J. Agric. Biol*., 20: 1167–1174

root inducing processes (Gelvin, 2009; Chandra, 2012). *Vir* genes are directly involved in T-DNA cleavage from the Ri plasmid, T-DNA processing, transferring and integration into the plant nuclei. Conversely, *chv* genes are not directly involved in the T-DNA transfer process. Instead, *chv* genes play important roles in signal transduction necessary for *Agrobacterium* pathogenicity (Subramoni *et al*., 2014; Nester, 2015). Chromosomal genes in *Agrobacterium* with pleiotropic functions also play important roles in *Agrobacterium*-mediated plant transformation (Gelvin, 2008). However, whether other genes in *Agrobacterium rhizogenes* chromosome are also linked to the invasion process remains elusive.

In 1987, the T-DNA region of the plasmid in *A. rhizogenes* strain K599 was identified and characterized (Combard *et al*., 1987). Using electron microscope to examine heteroduplexes, pRi2659 in K599 is confirmed to mainly share two homologous sequences of about 2.5 and 1.5 kb in size, respectively, bracketing a largely non-homologous central part about 5.5 kb long with Ri plasmids 1855 and 8196 (Brevet and Tempé, 1988). In 2001, the nucleotide sequence of cucumopine synthase encoded on pRi2659 was determined (Suzuki *et al*., 2001). Subsequently, the sequences of *rol* genes, *orf3*, *cus*, T-DNA, and pRi2659 in K599 were reported (Serino *et al*., 1994; Mankin *et al*., 2007; Wang *et al*., 2016) (GenBank accession number: Z29365; AJ271050; AB039860; EF433766; NC_010841; EU186381; CS354244). The draft genome sequence was also reported in 2016 (Valdes Franco *et al*., 2016). However, the whole genome sequence of K599 has yet to be determined.

Here, we selected *A. rhizogenes* strain K599 as our model to sequence and assemble its complete genome. The genetic information of K599 significantly advances our understanding of the transformation mechanism of T-DNA transfer and other crucial processes, and is of great usage to determine the limits of recalcitrant host species.

Materials and Methods

Growth Conditions and Genomic DNA Preparation

A. rhizogenes strain K599 was cultured on solid and in liquid YEM medium plus 50 mg/L streptomycin at 28^{o}C. For isolation of the genomic DNA, 1.5 g fresh bacterial cells from an overnight suspension culture were used and the extraction procedure was performed following the cetyl-trimethylammonium bromide (CTAB) method protocol (Del Sal *et al*., 1989). The purity and the concentration of the extracted DNA were measured with a UV spectrophotometer (TBS380, Promega, Madison, USA) while its integrity was evaluated electrophoretically in a 0.8% agarose gel.

Morphological Character and Phylogenetic Analysis

Confocal laser scanning microscope (Leica Microsystems GmbH, Wetzlar, Germany) was employed for imaging of K599. Phylogenetic analysis is based on 16S rDNA gene alignments and was conducted in MEGA7.0 (Kumar *et al*., 2016) with bootstrap value of 500 replicates.

Genome Sequencing and Assembly

A paired-end (2 × 251 bp) library with an insert size of 400 bp using TruSeq DNA PCR-Free Sample Preparation Kit (Illumina Inc.) and a standard library with an insert size of 10 Kb using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences of California, Inc., USA) were constructed, respectively. Using the next-generation sequencing (NGS) and the single molecule sequencing (SMS) (*i.e.* third generation sequencing (TGS)) technologies, the whole genome was sequenced using a combination of Illumina MiSeq and Pacbio RS II platforms at Shanghai Personal Biotechnology Co., Ltd. (China). Paired-end sequence reads were generated using the Illumina MiSeq. FASTQ sequence files were obtained using the Illumina Casava pipeline v1.8.3 (Illumina version1.8+). FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used to assess the quality resulting in 3,036,286 reads. Adapter- and quality-trimming was performed in Adapter Removal (version 2) (Lindgreen, 2012) and Quake (version 0.3) (Kelley *et al*., 2010). In total, 3,077,772 reads and 691,702,734 bp HQ data were obtained. Moreover, the data collected from the PacBio RSII instrument were processed and filtered using the SMRT analysis software suite. The Continuous Long Read data were filtered by Read-length (>50), Subread-length (>50) and Read quality (>0.75) resulting in 304,478 reads and 1,144,062,324 bp. Assembly was completed with the Hierarchical Genome Assembly Process algorithm implemented in the PacBio SMRT analysis software (Pacific Biosciences, Menlo Park, CA, USA). Using Newbler version 2.8 (Roche Co.), data corrected by Kmer were assembled to get contigs and scaffolds. Scaffolds were produced with The Hierarchical Genome Assembly Process (HGAP) from TGS data. The collinearity of contigs was analyzed by mummer software (http://www.tigr.org/software/mummer), and the contigs were reconfirmed to fill gaps. Finally, Pilon software was used to correct the results, and the complete sequence was obtained (Table 1).

Data from NGS and TGS were assembled to obtain the whole genome of 5,114,384 bp in size (GenBank accession number is NZ_CP019701, NZ_CP019702 and NZ_CP019703) with 359-fold overall coverage (135× and 224×, respectively) and a GC content of 59.75%, which includes one circular chromosome of 2,622,758 bp, one linear chromosome of 2,289,334 bp, and one Ri plasmid (pRi2659) of 202,292 bp (Table 1 and Table 2).

Genome Annotation

Glimmer 3.0 (Delcher *et al*., 1999) was used to predict protein-coding gene in genome, with the criteria of open reading frame (ORF) length longer than 110 bp, and the rest parameters default.

Table 1: Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	Paired-end (2×251 bp) Illumina library with inserts 400 bp and PacBio libraries with inserts of 10 Kb, respectively
MIGS 29	Sequencing platforms	Illumina Miseq and PacBio sequencer, respectively
MIGS 31.2	Fold coverage	135× and 224×, respectively
MIGS 30	Assemblers	Newbler (version 2.8) software for NGS sequences and Hierarchical Genome Assembly Process (HGAP) algorithm implemented in the PacBioSMRT Analysis software for TGS sequences, Last using mummer software and pilon software for contigs analysis
MIGS 32	Gene calling method	Glimmer 3.0 gene prediction
	Locus Tag	B0909
	GenBank ID	NZ_CP019701, NZ_CP019702, NZ_CP019703
	GenBank Date of Release	March 09, 2017
	GOLD ID	Not available
	BIOPROJECT	PRJNA373797
MIGS 13	Source Material Identifier	Not available
	Project relevance	agricultural

Table 2: Summary of genome: two chromosomes and one plasmid

Label	Size (Mb)	Topology	INSDC identifier	RefSeq ID
Chromosome 1	2.622758	Circular	PRJNA373797	NZ_CP019701
Chromosome 2	2.289334	Linear	PRJNA373797	NZ_CP019702
Plasmid	0.202292	Circular	PRJNA373797	NZ_CP019703

Table 3: Classification and general features of *Agrobacterium rhizogenes* strain K599 according to the MIGS recommendations (Field *et al*., 2008)

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS (Woese *et al*., 1990)
		Phylum *Proteobacteria*	TAS (Garrity *et al*., 2005)
		Class *Alphaproteobacteria*	TAS (Garrity *et al*., 2005)
		Order *Rhizobium*	TAS (Euzéby, 2006)
		Family *Rhizobiaceae*	TAS (Skerman *et al*., 1980)
		Genus *Agrobacterium*	TAS (Conn, 1942; Sawada *et al*., 1993; Young *et al*., 2001)
		Species *Agrobacterium rhizogenes*	TAS (Conn, 1942; Sawada *et al*., 1993; Young *et al*., 2001)
		Type strain: K599 (NCPPB2659)	
	Gram stain	Negative	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
	Cell shape	Rod	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
	Motility	Motile	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
	Sporulation	Not reported	NAS
	Temperature range	25-30°C	IDA
	Optimum temperature	28°C	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
	pH range; Optimum	5-9; 6.5-7.5	IDA
	Carbon source	sucrose, mannitol	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659; Du *et al*., 2015)
MIGS-6	Habitat	Soil and rhizosphere of plant	TAS (http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-6.3	Salinity	Up to 0.5% NaCl (w/v)	TDA
MIGS-22	Oxygen requirement	Aerobic or aerotolerant	TAS (http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-15	Biotic relationship	free-living or commensal	TAS(http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-14	Pathogenicity	Pathogenic: inducing hairy root	TAS (Combard *et al*., 1987; Combard and Baucher, 1988; http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-4	Geographic location	United Kingdom	TAS (http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-5	Sample collection	1974	TAS (http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659)
MIGS-4.1	Latitude	Not reported	NAS
MIGS-4.2	Longitude	Not reported	NAS
MIGS-4.4	Altitude	Not reported	NAS

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (*i.e*., a direct report exists in the literature); NAS: Non-traceable Author Statement (*i.e*., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project (Ashburner *et al*., 2000)

tRNAscan-SE (version 1.3.1) and RNAmmer (version 1.2) were used to predict tRNA genes (Lowe and Eddy, 1997) and rRNA genes (Lagesen *et al*., 2007) in the genome. Other ncRNAs were predicted using Rfam database (Burge *et al*., 2013). Signal peptides and transmembrane helices were predicted using Signal P 4.1 (Petersen *et al*., 2011) and

TMHMM (version 2.0c) (Chen *et al*., 2003), respectively. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated sequence (CAS) proteins constitute a putative prokaryotic RNA-interference-based immune system (Makarova *et al*., 2006; Sorek *et al*., 2008) protecting against bacteriophages or plasmids (Mojica *et al*., 2005; Smits *et al*., 2010). Using CRISPR recognition tool (CRT), direct repeats (DRs) and spacers in genome were predicted. CRISPRFinder was used for CRISPR identification.

Results

Classification and Features

K599 was isolated from cucumber (*Cucumis sativus* L.) in United Kingdom by Catton in 1974 (http://ncppb.fera.defra.gov.uk/furtherinfo.cfm?ncppb_no=2659). It is a Gram-negative bacterium featuring faint yellow colonies (Fig. 1A). Cells are rod shaped and are 2 μm or so in length (Fig. 1B). The strain is moderately fast-growing, forming 1 mm colonies within 1–2 days at 28°C. Phylogenetic analysis using 16S rDNA genes of other *Agrobacterium* strains and related members of the same family (*Microbacteriaceae*) and class (*Alphaproteobacteria*) is shown in Fig. 2. The general features of the strain are summarized in Table 3.

K599 infects a wide range of plant hosts to induce hairy roots (Xiang *et al*., 2016) (Fig. 3). Hairy roots induced by K599 can be directly used to study gene functions and functional genomics (Jian *et al*., 2009), providing a tool for understanding plant-microorganism and plant-plant interactions (Runo *et al*., 2012), for producing valuable secondary metabolites or compounds of interest (Tavizi *et al*., 2015), and for generating composite transgenic plant (non-transgenic shoots with transgenic roots) (Collier *et al*., 2005) and complete transgenic plants (Zhang *et al*., 2011).

Genome Properties

The complete genome of K599 is composed of one circular chromosome (2,622,758 bp), one linear chromosome (2,289,334 bp), and one circular plasmid (202,292 bp). The size of its genome is 5,114,384 bp with an average GC content of 59.75%. Among the 4889 predicted genes, 4716 were identified as protein coding genes. 4716 (96.46%) were assigned with putative functions and annotated as hypothetical proteins. 70 RNAs, which include 54 tRNAs, 4 5S rRNAs, 4 16S rRNAs, 4 23S rRNAs and 4 ncRNAs, and 103 pseudogenes were also identified.

The detailed information about K599 genome properties is summarized in Table 4 and genes assigned to COG (Galperin *et al*., 2015) functional categories are listed in Table 5. Sequences of the genome, predicted genes and non-encoding RNAs in GenBank format are mapped using CGView (Stothard and Wishart, 2005) (Fig. 4).

Fig. 1: Source organism *A. rhizogenes* strain K599. **A**: The morphology of its colony after 36 h growing on YEM solid medium plus 50 mg/L streptomycin; **B**: Confocal laser scanning microscope imaging of K599

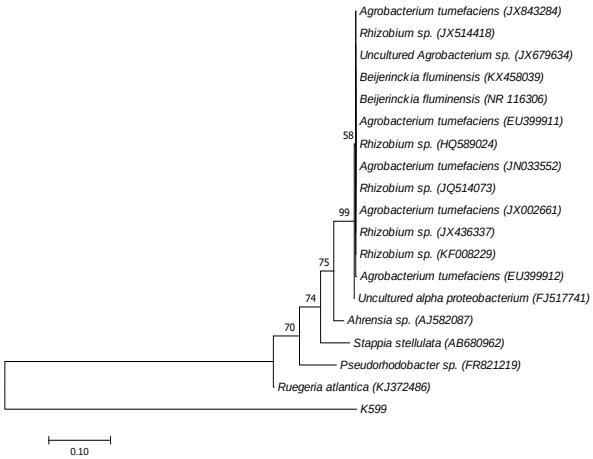

Fig. 2: The phylogenetic tree showing the relationship of *A. rhizogenes* strain K599 with other strains of *Agrobacterium*, *Ahrensia*, *Pseudorhodobacter*, *Ruegeria* and *Stappia*. GenBank accession numbers are shown in parenthesis

Insights from the Genome Sequence

Using PHAge Search Tool (PHAST) (Zhou *et al*., 2011), 5 prophage regions have been identified in the genome, of which 2 regions are intact (in chromosome 1), 3 regions are incomplete (one in chromosome 1 and two in

Table 4: Genome statistics

Attribute	Value	% of Total
Genome size (bp)	5,114,384	100.00
DNA coding (bp)	4,487,487	87.74
DNA G+C (bp)	1,686,886	59.75
DNA scaffolds	3	100
Total genes	4889	100
Protein coding genes	4716	96.46
RNA genes	70	1.43
Pseudo genes	103	2.11
Genes in internal clusters	NA	-
Genes with function prediction	4716	96.46
Genes assigned to COGs	2306	47.17
Genes with Pfam domains	2117	43.3
Genes with signal peptides	427	8.73
Genes with transmembrane helices	1229	25.14
CRISPR repeats	1	0.02

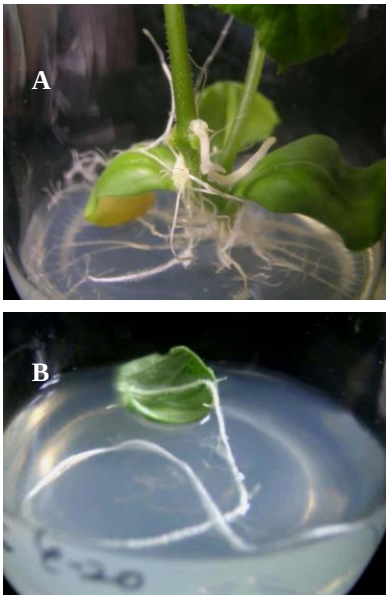

Fig. 3: Hairy roots induced from cucumber infected by *A. rhizogenes* strain K599. A: Hairy roots from epicotyls *in vivo*; B: Hairy roots from cotyledons *in vitro*

chromosome 2, respectively). Prophages are essentially dormant phages that are only replicated through bacterial DNA replication and cell division. Defective or cryptic prophages are abundant in many bacterial genomes and they can carry a number of genes that may be beneficial to the host. Bacteria containing a prophage are called lysogens because their prophage is in the lysogenic cycle, in which the viral (esp. the lytic) genes are not expressed. Upon damage to the host cell DNA or other physiological cues, the prophage may be induced to excise itself from the bacterial genome. Bacterial genomes can contain a significant proportion (>20%) of functional and non-functional bacteriophage genes (Casjens, 2003). The presence of prophage sequences may also allow some bacteria to acquire antibiotic resistance, to survive in new environmental niches, to improve adhesion or to become

Table 5: Number of genes associated with general COG functional categories

Code	Value	%age	Description
J	133	4.95	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	187	6.95	Transcription
L	82	3.05	Replication, recombination and repair
B	1	0.04	Chromatin structure and dynamics
D	11	0.41	Cell cycle control, Cell division, chromosome partitioning
V	28	1.04	Defense mechanisms
T	108	4.02	Signal transduction mechanisms
M	119	4.43	Cell wall/membrane biogenesis
N	47	1.75	Cell motility
U	43	1.60	Intracellular trafficking and secretion
O	100	3.72	Posttranslational modification, protein turnover, chaperones
C	105	3.90	Energy production and conversion
G	135	5.02	Carbohydrate transport and metabolism
E	241	8.96	Amino acid transport and metabolism
F	68	2.53	Nucleotide transport and metabolism
H	64	2.38	Coenzyme transport and metabolism
I	72	2.68	Lipid transport and metabolism
P	154	5.73	Inorganic ion transport and metabolism
Q	52	1.93	Secondary metabolites biosynthesis, transport and catabolism
R	316	11.75	General function prediction only
S	240	8.93	Function unknown
-	684	25.44	Not in COGs

The total is based on the total number of protein coding genes in the genome

pathogenic. Because bacterial genome fragments can also be carried by phage particles, the lytic process is thought to be an important vehicle for horizontal gene transfer (Casjens, 2003; Coates and Hu, 2011).

We conducted BLAST to VFDB database and set the minimum blastn percentage identity at 45% and the ratio of the length of sequence used alignment to its full-length is not less than 70%. Thirteen virulent genes in two chromosomes (ten in chromosome 1 and three in chromosome 2, respectively) were predicted in genome based on VFDB. Most bacterial virulence factors were originally thought to be associated with pathogens. We further identified genes most likely to be involved in virulence of infection to plant cell. The virulence factor database (VFDB) is an integrated and comprehensive online resource for curating virulence factors of bacterial pathogens (Chen *et al.*, 2012).

Discussion

Combining whole genome shotgun strategy using next-generation sequencing (NGS) and the third-generation sequencing (TGS) technologies, we *do novo* assembled the complete genome of K599. Using Illumina Miseq and Pacbio RS II, we obtained 3,036,286 high-quality filtered sequence reads of 691,702,734 bp, and 304,478 sequence reads of 1,144,062,324 bp, respectively. The whole genome size is 5,114,385 bp, which consists one circular chromosome (2,622,758 bp), one linear chromosome (2,289,334 bp), and one Ri plasmid (*i.e.*, pRi2659) (202,292 bp).

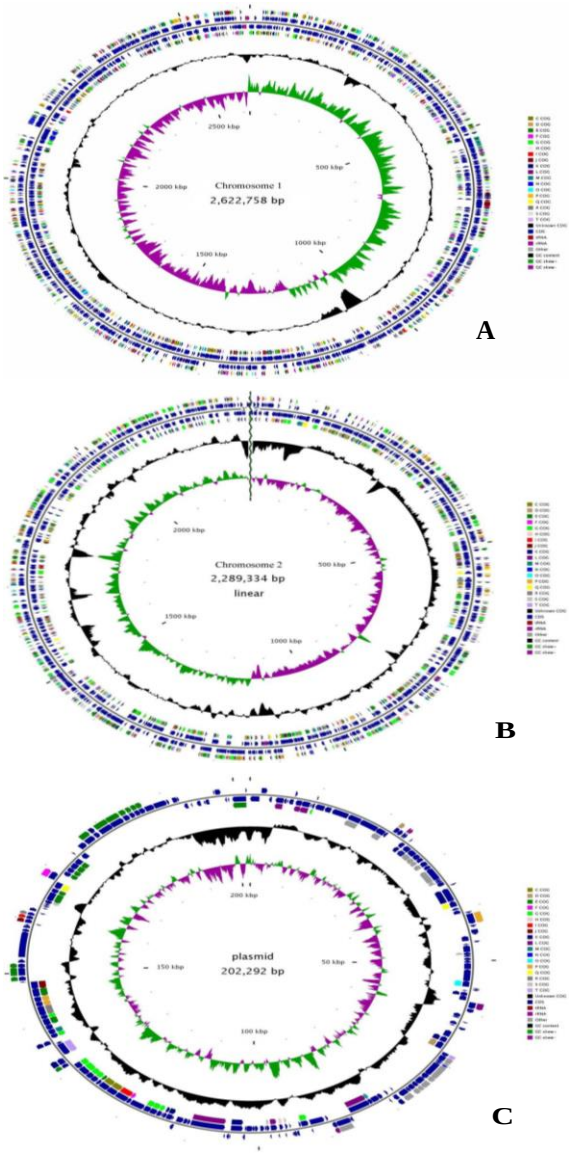

Fig. 4: Genome maps of *A. rhizogenes* strain K599. **A**: Graphic representation of chromosome 1 (circular); **B**: Graphic representation of chromosome 2 (linear); **C**: Graphic representation of Ri plasmid pRi2659 (circular). Circles (from inside to outside): first, scale bar in kilobases; second, G+C skew (G-C/G+C): values>0 (green), values<0 (purple); third, G+C content: above median G+C content (outside), less than or equal to the median (inside); fourth and seventh, predicted coding sequences on the plus and minus strands, respectively (colors according to COGs); fifth and sixth, positions of coding sequences

The complete genome sequence has an overall G+C content of 59.75%. The first version of annotation includes 12 rRNA genes (four 5S rRNAs, four 16S rRNAs and four 23S rRNAs, respectively), 54 tRNA genes, 4 other ncRNAs and 4716 protein-coding genes. GenBank accession numbers for the genome sequence of K599 are NZ_CP019701 (chromosome 1), NZ_CP019702 (chromosome 2) and NZ_CP019703 (Ri plasmid of pRi2659).

The size of our complete genome sequences is 162,963 bp shorter than that of the reported draft genome sequence (Valdes Franco *et al*., 2016). We speculate that the draft genome sequence using NGS contains severally artefactual duplication, including errors in merging some adjacent plasmid-clone sequences by the assembly program. By combining the NGS with TGS technologies, we eliminated these artefacts in our assembly process.

Although the whole genomes of three *Agrobacterium tumefaciens* strains were reported (https://www.ncbi.nlm.nih.gov/genome/genomes/177?), as to the best of our knowledge, this is the first report of the whole genome of *Agrobacterium rhizogenes.* The accuracy and completeness of the current genome sequence of K599 may facilitate developing *Agrobacterium rhizogenes* as tools for plant transformation.

Acknowledgements

The work was supported by grants from the National Key Research and Development Project of China (Grant No. 2016YFC0503100), the Project of Science and Technology of Zhejiang Province for Commissioner (T Xiang). We acknowledge Shanghai Personal Biotechnology Co. Ltd. for genome sequence.

References

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock, 2000. Gene ontology: tool for the unification of biology. *Nat. Genet*., 25: 25–29

Banta, L.M. and M. Montenegro, 2008. Agrobacterium and Plant Biotechnology. *In: Agrobacterium: From Biology to Biotechnology*, pp: 73–148. Tzfira, T. and V. Citovsky (Eds.). Springer Publishing Company, New York, USA

Brevet, J. and J. Tempé, 1988. Homology mapping of T-DNA regions on three *Agrobacterium rhizogenes* Ri plasmids by electron microscope heteroduplex studies. *Plasmid*, 19: 75–83

Burge, S.W., J. Daub, R. Eberhardt, J. Tate, L. Barquist, E.P. Nawrocki, S.R. Eddy and P.P. Gardner, 2013. Bateman A: Rfam 11.0: 10 years of RNA families. *Nucl. Acids Res.*, 41: 226–232

Casjens, S., 2003. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.*, 49: 277–300

Chandra, S., 2012. Natural plant genetic engineer *Agrobacterium rhizogenes*: role of T-DNA in plant secondary metabolism. *Biotechnol. Lett.*, 34: 407–415

Chen, L., Z. Xiong, L. Sun, J. Yang and Q. Jin, 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucl. Acids Res.*, 40: 641–645

Chen, Y., P. Yu, J. Luo and Y. Jiang, 2003. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm. Genom.*, 14: 859–865

Chilton, M.D., D.A. Tepfer, A. Petit, C. David, F. Casse-Delbart and J. Tempe, 1982. *Agrobacterium rhizogenes* insert T-DNA into the genomes of the host plant root cells. *Nature*, 295: 432–434

Cho, H.J., S.K. Farrand, G.R. Noel and J.M. Widholm, 2000. High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. *Planta*, 210: 195–204

Coates, A.R. and Y. Hu, 2011. Novel approaches to developing new antibiotics for bacterial infections. *Brit. J. Pharmacol.*, 152: 1147–1154

Collier, R., B. Fuchs, N. Walter, W.K. Lutke and C.G. Taylor, 2005. *Ex vitro* composite plants: an inexpensive, rapid method for root biology. *Plant J.*, 43: 449–457

Combard, A. and M.F. Baucher, 1988. A common organization of the T-DNA genes expressed in plant hairy roots induced by different plasmids of *Agrobacterium rhizogenes*. *Plant Mol. Biol*., 10: 499–509

Combard, A., J. Brevet, D. Borowski, K. Cam and J. Tempé, 1987. Physical map of the T-DNA region of *Agrobacterium rhizogenes* strain NCPPB2659. *Plasmid*, 18: 70–75

Conn, H.J., 1942. Validity of the Genus *Alcaligenes*. *J. Bacteriol*., 44: 353–360

Delcher, A.L., D. Harmon, S. Kasif, O. White and S.L. Salzberg, 1999. Improved microbial gene identification with GLIMMER. *Nucl. Acids Res.*, 27: 4636–4641

Del Sal, G., G. Manfioletti and C. Schneider, 1989. The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *Biotechniques*, 7: 514–520

Du, S., T. Xiang, Y. Song, L. Huang, Y. Sun and Y. Han, 2015. Transgenic hairy roots of *Tetrastigma hemsleyanum*: induction, propagation, genetic characteristics and medicinal components. *Plant Cell Tissue Org. Cult.*, 122: 373–382

Euzéby, J., 2006. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int. J. Syst. Evol. Microbiol*., 56: 1–6

Field, D., G. Garrity, T. Gray, N. Morrison, J. Selengut, P. Sterk, T. Tatusova, N. Thomson, M.J. Allen and S.V. Angiuoli, 2008. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat. Biotechnol*., 26: 541–547

Galperin, M.Y., K.S. Makarova, Y.I. Wolf and E.V. Koonin, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucl. Acids Res.*, 43: 261–269

Garrity, G.M., J.A. Bell and T. Lilburn, 2005. Class I. *Alphaproteobacteria* class. nov. *In: Bergey's Manual of Systematic Bacteriology*, 2nd edition, Volume 2, Part C, p: 1. Garrity, G.M., D.J. Brenner, N.R. Krieg and J.T. Staley (Eds.). Springer Publishing Company, New York, USA

Gelvin, S.B., 2009. *Agrobacterium* in the genomics age. *Plant Physiol.*, 150: 1665–1676

Gelvin, S.B., 2008. Function of host proteins in the Agrobacterium-mediated plant transformation process. *In: Agrobacterium: From Biology to Biotechnology*, pp: 483–522. Tzfira, T. and V. Citovsky (Eds.). Springer Publishing Company, New York

Georgiev, M.I., E. Agostini, J. Ludwig-Müller and J. Xu, 2012. Genetically transformed roots: from plant disease to biotechnological resource. *Trends Biotechnol*., 30: 528–537

Guillon, S., J. Tremouillaux-Guiller, P.K. Pati, M. Rideau and P. Gantet, 2006. Hairy root research: Recent scenario and exciting prospects. *Curr. Opin. Plant Biol*., 9: 341–346

Jian, B., W. Hou, C. Wu, B. Liu, W. Liu, S. Song, Y. Bi and T. Han, 2009. *Agrobacterium rhizogenes*-mediated transformation of Superroot-derived Lotus corniculatus plants: a valuable tool for functional genomics. *BMC Plant Biol*., 9: 78

Kelley, D.R., M.C. Schatz and S.L. Salzberg, 2010. Quake: quality-aware detection and correction of sequencing errors. *Genom. Biol.*, 11: 116

Kumar, S., G. Stecher and K. Tamura, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*., 33: 1870–1874

Lagesen, K., P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes and D.W. Ussery, 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl. Acids Res.*, 35: 3100–3108

Lindgreen, S., 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes*, 5: 337

Lowe, T.M. and S.R. Eddy, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res*., 25: 955–964

Makarova, K.S., N.V. Grishin, S.A. Shabalina, Y.I. Wolf and E.V. Koonin, 2006. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*., 1: 7

Mankin, S.L., D.S. Hill, P.M. Olhoft, E. Toren, A.R. Wenck, L. Nea, L. Xing, J.A. Brown, H. Fu, L. Ireland, H. Jia, H. Hillebrand, T. Jones and H-S. Song, 2007. Disarming and sequencing of *Agrobacterium rhizogenes* strain K599 (NCPPB2659) plasmid pRi2659. *Vitro Cell Dev. Biol. Plant*, 43: 521–535

Mojica, F.J., C. Díez-Villaseñor, J. García-Martínez and E. Soria, 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol*., 60: 174–182

Nester, E.W., 2015. Agrobacterium: nature's genetic engineer. *Front. Plant Sci*., 5: 730

Petersen, T.N., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8: 785–786

Runo, S., S. Macharia, A. Alakonya, J. Machuka, N. Sinha and J. Scholes, 2012. Striga parasitizes transgenic hairy roots of *Zea mays* and provides a tool for studying plant-plant interactions. *Plant Methods*, 8: 1–11

Savka, M.A., B. Ravillion, G.R. Noel and S.K. Farrand, 1990. Induction of hairy roots on cultivated soybean genotypes and their use to propagate the soybean cyst nematode. *Phytopathology*, 80: 503–508

Sawada, H., H. Ieki, H. Oyaizu and S. Matsumoto, 1993. Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *Int. J. Syst. Bacteriol.*, 43: 694–702

Serino, G., D. Clerot, J. Brevet, P. Costantino and M. Cardarelli, 1994. *rol* genes of *Agrobacterium rhizogenes* cucumopine strain: sequence, effects and pattern of expression. *Plant Mol. Biol.*, 26: 415–422

Smits, T.H., F. Rezzonico, T. Kamber, J. Blom, A. Goesmann, J.E. Frey and B. Duffy, 2010. Complete genome sequence of the fire blight pathogen *Erwinia amylovora* CFBP 1430 and comparison to other *Erwinia* spp. *Mol. Plant Microbiol. Interact*., 23: 384–393

Skerman, V.B.D., V. McGowan and P.H.A. Sneath, 1980. Approved Lists of Bacterial Names. *Int. J. Syst. Evol. Microbiol*., 30: 225–420

Sorek, R., V. Kunin and P. Hugenholtz, 2008. CRISPR-a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol*., 6: 181–186

Stothard, P. and D.S. Wishart, 2005. Circular genome visualization and exploration using CGView. *Bioinformatics*, 21: 537–539

Subramoni, S., N. Nathoo, E. Klimov and Z.C. Yuan, 2014. *Agrobacterium tumefaciens* responses to plant-derived signaling molecules. *Front. Plant Sci*., 5: 322

Suzuki, K., N. Tanaka, H. Kamada and I. Yamashita, 2001. Mikimopine synthase (mis) gene on pRi1724. *Gene*, 263: 49–58

Tavizi, A., M.J. Javaran, A. Moieni, M. Mohammadi-Dehcheshmeh, M. Mohebodini and E. Ebrahimie, 2015. Root and shoot parts of strawberry: factories for production of functional human pro-insulin. *Mol. Biol. Rep*., 42: 1013–1023

Valdes Franco, J.A., R. Collier, Y. Wang, N. Huo, Y. Gu, R. Thilmony and J.G. Thomson, 2016. Draft genome sequence of *Agrobacterium rhizogenes* Strain NCPPB2659. *Genom. Announc.*, 4: e00746–16

Wang, S., Y. Song, T. Xiang, P. Wu, T. Zhang, D. Wu, S. Zhou and Y. Li, 2016. Transgenesis of *Agrobacterium rhizogenes* K599 *orf3* into plant alters plant phenotype to dwarf and branch. *Plant Cell Tiss. Org. Cult*., 127: 207–215

White, F.F. and E.W. Nester, 1980. Hairy root: plasmid encodes virulence traits in *Agrobacterium rhizogenes*. *J. Bacteriol*., 141: 1134–1141

Woese, C.R., O. Kandler and M. L. Wheelis, 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA*, 87: 4576–4579

Xiang, T., S. Wang, P. Wu, Y. Li, T. Zhang, D. Wu and S. Zhou, 2016. Cucumopine type *Agrobacterium rhizogenes* K599 (NCPPB2659) T-DNA mediated plant transformation and its application. *Bangl. J. Bot.*, 45: 935–945

Young, J.M., L.D. Kuykendall, E. Martínez-Romero, A. Kerr and H. Sawada, 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*. *Int. J. Syst. Evol. Microbiol*., 51: 89–103

Zhang, D., T. Xiang, P. Li and L. Bao, 2011. Transgenic plants of *Petunia hybrida* harboring the *CYP2E1* gene efficiently remove benzene and toluene pollutants and improve resistance to formaldehyde. *Genet. Mol. Biol*., 34: 634–639

Zhou, Y., Y. Liang, K.H. Lynch, J.J. Dennis and D.S. Wishart, 2011. PHAST: a fast phage search tool. *Nucl. Acids Res*., 39: 347–352

(Received 22 January 2018; Accepted 26 January 2018)