*De novo* genome assembly of plant-associated *Rhodococcus qingshengii* (RL1) isolated from *Eruca sativa* Mill. showing plant growth promoting properties

Theresa Kuhl[a#], Marius Felder[b*], Thomas Nussbaumer[a,d], Doreen Fischer[c], Susanne Kublik[c], Soumitra Paul Chowdhury[a], Michael Schloter[c], Michael Rothballer[a]

**Abstract** *Rhodococcus qingshengii* RL1 was isolated from surface-sterilized leaves of *Eruca sativa* Mill. and shows plant growth promoting (PGP) properties. The *de novo* genome assembly consists of one chromosome with 6.253.838 bp and two plasmids with 144.038 bp and 448.745 bp. Many genes could be identified reflecting its PGP potential.

The genus *Rhodococcus* belongs to the phylum *Actinobacteria* and includes aerobic, Gram-positive, nonsporulating bacteria isolated from a broad variety of environments (1–3). Some of these bacteria have large genomes (> 5 Mb) with high GC content. Their ability to degrade a large spectrum of environmentally problematic compounds (2, 4) or perform quorum quenching (5) makes them suitable for bioremediation or agricultural applications.

RL1 was isolated from surface-sterilized leaves of *Eruca sativa* Mill. The *de novo* assembly of the RL1 genome showed highest similarity of > 99% sequence identity with over 90% of the *Rhodococcus qingshengii* djl-6$^T$ genome, and the 16S rDNA of RL1 was 99,9% identical with djl-6$^T$. In a 16S rDNA based phylogenetic neighbor-joining tree (6) calculated with ARB (7) RL1 was placed within a cluster (bootstrap support 99%) consisting of *Rhodococcus erythropolis*$^T$, *Rhodococcus qingshengii* djl-6$^T$, *Rhodococcus degradans* CCM 4446$^T$ and *Rhodococcus baikonurensis*$^T$ (Fig.1). Further phylogenetic analysis of the *gyrB* gene verified the taxonomic classification of RL1 as *Rhodococcus qingshengii* (data not shown). RL1 grows in Tryptic Soy Broth at 28°C and is tolerant of NaCl up to a concentration of 7.5%. The isolate shows PGP traits (indole acetic acid production (IAA), siderophore production, phosphate solubilization, 1-aminocyclopropane-1-carboxylate (ACC) deaminase activity, root elongation) and is able to colonize the root surface of its host (*E. sativa*). For sequencing

genomic DNA from RL1 was isolated via standard phenol-chloroform extraction with previous lysis with Ampicillin 600µg/mL for 3h. For the PacBio Sequel System, the library was prepared with SMRTbell™ Template Prep Kit 1.0-SPv3 and SMRTbell Barcoded Adapt Complete Prep-96. PacBio sequencing was performed with Sequel Sequencing Kit 2.0 (8 rxn) and SMRT Cell 1M v2 Tray. For Illumina MiSeq sequencing, the library was prepared using the TruSeq DNA PCR-free library preparation kit (Illumina, San Diego, CA, USA). Genomic DNA was fragmented applying the Covaris E220 system according to the manufacturer's protocol for a 550 bp average insert size and sequenced using MiSeq Reagent Kit v3 (600-cycle) (Illumina).

A total of 376,794 PacBio long-reads (209x coverage) and a total of 1,068,580 Illumina short-reads (49x coverage), quality checked with FastQC (8), were included in the *de novo* assembly of the RL1 genome using the hybrid assembler MaSuRCA (9). Sequence assembly produced three contigs representing one chromosome and two plasmids (chromosome: 6.253.838 bp, plasmid1: 144.038 bp, plasmid2: 448.745 bp) with 62.4% GC content. The chromosome and plasmid 1 were circularized with Circlator version 1.5.5 (10). 6.652 coding sequences were predicted with Rapid Annotations using Subsystems Technology (RAST) (11) and gene clusters were identified with antiSMASH 4.2.0 (12) and Plant-bacteria Interaction Factors Resource (PIFAR) (13). All tools were used with default settings. Many identified gene clusters were associated with traits relevant for (beneficial) microbe-plant interactions: siderophore production, IAA production, osmoregulation (Ectoine), glucosinolate metabolism (ß-Glucosidase, *msrB*), quorum quenching (*qsdA*), antibiotics production, biofilm formation, lipopolysaccharides production, multidrug resistance, microbe-associated molecular patterns (MAMPs), heavy metal tolerance and reactive oxygen species resistance.

**Accession Number**: This whole-genome sequencing project was deposited in GenBank under accession no. CP042915, CP042916, CP042917.

**Figure legend**:

Figure 1: Neighbor-joining phylogenetic tree based on 16S rDNA sequences showing the phylogenetic relationship between *Rhodococcus qingshengii* RL1 and other members of the genus *Rhodococcus* and the family *Nocardiacea*. Bootstrap values (%) for 1000 resamplings are given at the nodes. Two versions of the 16S rRNA gene in the RL1 genome are included, differing at position 1074 (A or C) of the complete 16S rRNA gene.

**References**


1. Xu JL, He J, Wang ZC, Wang K, Li WJ, Tang SK, Li SP. 2007. Rhodococcus qingshengii sp. nov., a carbendazim-degrading bacterium. Int J Syst Evol Microbiol 57:2754–2757.

2. Lincoln SA, Hamilton TL, Valladares Juárez AG, Schedler M, Macalady JL, Müller R, Freeman KH. 2015. Draft Genome Sequence of the Piezotolerant and Crude Oil-Degrading Bacterium *Rhodococcus qingshengii* Strain TUHH-12. Genome Announc 3:e00268-15.

3. Monu Bala,a Shailesh Kumar,b Gajendra Pal Singh Raghava b SM. 2013. Draft Genome Sequence of Rhodococcus qingshengii Strain BKS 20-40 Monu. Genome Announc 1.

4. Kamble AL, Banoth L, Meena VS, Singh A, Chisti Y, Banerjee UC. 2013. Nitrile hydratase of Rhodococcus erythropolis: characterization of the enzyme and the use of whole cells for biotransformation of nitriles. 3 Biotech 3:319–330.

5. Rückert C, Birmes FS, Müller C, Niewerth H, Winkler A, Fetzner S, Kalinowski J. 2015. Complete genome sequence of Rhodococcus erythropolis BG43 (DSM 46869), a degrader of Pseudomonas aeruginosa quorum sensing signal molecules. J Biotechnol 211:99–100.

6. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

7. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A, Buchner A, Lai T, Steppi S, Jacob G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßbmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH. 2004. ARB: A software environment for sequence data. Nucleic Acids Res 32:1363–1371.

8. Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data.

9. Zimin A V., Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677.

10. Hunt M, Silva N De, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: Automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:1–10.

11. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:206–214.

12. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, De Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36–W41.

13. Martínez-García PM, López-Solanilla E, Ramos C, Rodríguez-Palenzuela P. 2016.

Prediction of bacterial associations with plants using a supervised machine-learning approach. Environ Microbiol 18:4847–4861.