Comparative genomics reveals diverse capsular polysaccharide synthesis gene clusters in emerging *Raoultella planticola*

Yao-Ting Huang¹, Wei-Yao Chuang¹, Bing-Ching Ho², Zong-Yen Wu²,³, Rita C Kuo³, Mengwei Ko³, Po-Yu Liu⁶,⁷,⁸,⁹,⁺

¹National Chung Cheng University, Department of Computer Science and Information Engineering, Chia-Yi, Taiwan
²National Taiwan University Hospital, Department of Clinical Laboratory Sciences and Medical Biotechnology, Taipei, Taiwan
³US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA
⁴National Chung Hsing University, Department of Veterinary Medicine, Taichung, Taiwan
⁵University of California, Division of Oral Biology and Oral Medicine, School of Dentistry and Medicine, The Jane and Jerry Weintraub Center for Reconstructive Biotechnology, Los Angeles, CA, USA
⁶Shu-Zen Junior College of Medicine and Management, Department of Nursing, Kaohsiung City, Taiwan
⁷National Chung Hsing University, College of Life Sciences, Rong Hsing Research Center for Translational Medicine, Taichung, Taiwan
⁸National Chung Hsing University, PhD Program in Translational Medicine, Taichung, Taiwan
⁹Taichung Veterans General Hospital, Department of Internal Medicine, Division of Infectious Diseases, Taichung, Taiwan

*Raoultella planticola* is an emerging zoonotic pathogen that is associated with rare but life-threatening cases of bacteremia, biliary tract infections, and urinary tract infections. Moreover, increasing antimicrobial resistance in the organism poses a potential threat to public health. In spite of its importance as a human pathogen, the genome of *Raoultella planticola* remains largely unexplored and little is known about its virulence factors. Although lipopolysaccharides have been detected in *R. planticola* and implicated in the virulence in earlier studies, the genetic background is unknown. Here, we report the complete genome and comparative analysis of the multidrug-resistant clinical isolate GODA. The complete genome sequence of *R. planticola* GODA was sequenced using single-molecule real-time DNA sequencing. Comparative genomic analysis reveals distinct capsular polysaccharide synthesis gene clusters in *R. planticola* GODA. In addition, we found bla₄₃₀⁵ and multiple transporters related to multidrug resistance. The availability of genomic data in open databases of this emerging zoonotic pathogen, in tandem with our comparative study, provides better understanding of *R. planticola* and the basis for future work.

Key words: *Raoultella planticola* - carbapenem resistance - capsular polysaccharide
was performed using the PacBio sequencing platform (Pacific Biosciences). Sequence runs of three single-molecule real-time (SMRT) cells were performed on the PacBio RS II sequencer with a 120-minute movie time/SMRT cell. SMRT Analysis portal version 2.1 was used for read filtering and adapter trimming, with default parameters, and post-filtered data of 1.2Gb (around 214X coverage) with an average read length of 6 kb were used for subsequent assembly.

The post-filtered reads were de novo assembled by Canu (v1.4) and converted into circular form via Circulator. These long reads were assembled and circularized into a complete circular genome (~5.6Mbp). Meanwhile, three additional plasmids were also reconstructed. The guanine-cytosine (GC) content of the GODA genome was 55.4%, which was similar with other related strains. Protein-coding genes in the genome and plasmids were annotated using NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP). Functional classification of annotated genes was carried out by RPSBLAST v. 2.2.15 in conjunction with the COGs (Clusters of Orthologous Groups of proteins) database. A total of 5,461 genes were identified, including 25 rRNA genes, and 83 tRNA genes (Table I).

We further constructed a pan-genome dataset using whole genome sequence of GODA and 7 publicly available whole genome sequences of R. planticola strains (Table I). We considered each gene to be strain-specific if it was present only in one strain and absent in all other strains. Furthermore, the genes shared by all strains were considered to be pan-genomic core genes. Fig. 1 shows orthologous genes shared among strains and depicts the position and color-coded function of the R. planticola GODA-specific genes. The numbers of orthologous and strain-specific unique genes are shown in the Venn diagrams (Fig. 2A). As presented in the figure, the pan genome of R. planticola revealed 4,382 core genes shared across all strains, whereas 147 genes were specific to R. planticola GODA. Functional analysis of GODA-specific genes revealed that, in addition to hypothetical proteins, a relative abundance of these gene are involved in replication and repair, followed by cell wall/membrane/envelop biogenesis (Fig. 2B). The Average Nucleotide Identity (ANI) was calculated based on a modified algorithm and revealed that R. planticola GODA is closely related to ATCC 33531, FDAARGOS_64, and CHB in terms of nucleotide sequences (ANI > 98%) (Fig. 3).

Virulence genes in the GODA genome were identified using the virulence factor database (VFDB). The identified virulence genes, which were also GODA-specific genes, were considered to be putative GODA-specific virulence factors.

The polysaccharide capsule is considered a major virulence factor of R. planticola (formerly named Klebsiella planticola). Previous study in Klebsiella spp. suggests the wzx is a common component in the capsular polysaccharide biosynthesis pathway. Our comparative genomics also revealed the presence of wzx flippase in the GODA genome, but this was lost in the environmental strains. Further investigation of its upstream and downstream genes revealed the entire capsular polysaccharide synthesis (cps) gene cluster (Fig. 4). Our findings provide the first genetic background of the cps gene clusters in R. planticola.

We further compared the cps clusters of environmental/clinical isolated strains and two distant-related Klebsiella strains (Fig. 4). Three highly conserved genes: galF, gnd and ugd were well-preserved across all strains analyzed, whereas the gene composition in between was often variable. A similar context has been noted in Klebsiella strains. The inter-species variability (R. planticola vs Klebsiella strains) was relatively higher than the intra-species variability. The cps structure of two clinical isolates, GODA and FDAARGOS_64, were highly similar, implying both strains may express identical virulence factors. While wzx was commonly found in Klebsiella spp., it was lost in all environmental isolated strains of R. planticola in this study.

Genetic context analysis of the capsular polysaccharide synthesis gene cluster of GODA showed that wzx was located between a gene encoding UTP-glucose-1-phosphate uridylyltransferase and a 6-phosphogluconate dehydrogenase. A similar observation has been made in several capsular polysaccharide synthesis gene clusters of Klebsiella spp. Capsular polysaccharide is a major virulence factor of Klebsiella spp. and genetic

### TABLE I

Features of Raoultella planticola strains in the study

| Strain      | Site of isolation | Country of origin | Genome assembly status | Genome size (bp) | GC content (%) | CDSs (pseudo genes) | rRNA operons | tRNAs |
|-------------|------------------|-------------------|------------------------|------------------|----------------|---------------------|-------------|-------|
| GODA        | Human            | Taiwan            | Complete               | 5,592,163        | 55.4           | 5,461(703)          | 25          | 83    |
| ATCC 33531  | Radish root      | Unknown           | Complete               | 5,668,028        | 55.8           | 5,363(193)          | 5           | 67    |
| CHB         | River            | USA               | Contig                 | 5,780,876        | 55.4           | 5,501(210)          | 24          | 77    |
| FDAARGOS_64 | Human            | USA               | Contig                 | 5,823,731        | 55.6           | 5,541(312)          | 25          | 86    |
| 1175_2058   | Human            | USA               | Contig                 | 5,750,464        | 55.7           | 5,486(96)           | 19          | 71    |
| 626_SENT    | Human            | USA               | Scaffold               | 5,735,751        | 55.5           | 5,544(233)          | 3           | 28    |
| INSali127   | Vegetable        | Portugal           | Scaffold               | 6,011,051        | 55.5           | 5,843(211)          | 5           | 72    |
| INSali133   | Vegetable        | Portugal           | Scaffold               | 6,011,836        | 55.5           | 5,840(220)          | 5           | 74    |

CDSs: coding sequences; GC: guanine-cytosine.
Fig. 1: circular genomes representation map and genome comparison of *Raoultella planticola* (GODA, 1175_2058, 626_SENT, ATCC 33531, CHB, FDAARGOS_64, INSali127, INSali133). Predicted coding sequences (CDSs) are assigned various colors with respect to cellular functions. Circles show from the outermost to the innermost: (1) DNA coordinates; (2, 3). Function-based color-coded mapping of the CDSs predicted on the forward and reverse strands of the *R. planticola* GODA genome, respectively; (4) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* 1175_2058; (5) *R. planticola* GODA-specific CDSs, compared with *R. planticola* 1175_2058; (6) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* 626_SENT; (7) *R. planticola* GODA-specific CDSs, compared with *R. planticola* 626_SENT; (8) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* ATCC 33531; (9) *R. planticola* GODA-specific CDSs, compared with *R. planticola* ATCC 33531; (10) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* CHB; (11) *R. planticola* GODA-specific CDSs, compared with *R. planticola* CHB; (12) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* FDAARGOS_64; (13) *R. planticola* GODA-specific CDSs, compared with *R. planticola* FDAARGOS_64; (14) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* INSali127; (15) *R. planticola* GODA-specific CDSs, compared with *R. planticola* INSali127; (16) Orthologous CDSs shared between *R. planticola* GODA and *R. planticola* INSali133; (17) *R. planticola* GODA-specific CDSs, compared with *R. planticola* INSali133; (18) GC plot with regions above and below average in green and violet; (19) GC skew showing regions above and below average in yellow and light blue. This figure was plotted in Scalable Vector Graphics format via an in-house script, which calculates the radius and ribbon width according to the BLAST alignments and adds colors by COG classification of all orthogonal genes.
structures of the capsular polysaccharide synthesis gene cluster in *Klebsiella* spp. have been well studied. Generally, *galF* at the 5’ end of the capsular polysaccharide regions and *gnd* and *ugd* at the 3’ end are highly conserved among different *Klebsiella*. The same context was identified in GODA. We also predicted genes encoding proteins necessary for capsular polysaccharide translocation and processing at the cell surface (*wza*, *wzb*, *wzc*, and *wzi*) and genes encoding glycosyltransferase.

The resistome in GODA was annotated using the Resistance Gene Identifier from the Comprehensive Antibiotic Resistance Database (CARD)\(^{(17)}\) and IMG database\(^{(18)}\). GODA showed the presence of *bla*\(_{TEM-57}\) (Table II), an extended-spectrum β-lactamase conferring resistance against β-lactam antibiotics such as penicillins and cephalosporins.\(^{(19)}\) GODA was also equipped with a number of efflux systems. GODA contains homologs of multidrug and toxic compound extrusion (MATE) family (*mdtK*), resistance-nodulation-cell division (RND) family (*mdtABC, ogxAB, acrAB*), ATP (adenosine triphosphate)-binding cassette (ABC) superfamily (*yojI, msbD*), and major facilitator superfamily (MFS).

Fig. 2: comparison of the gene contents of the *Raoultella planticola*. (A) Venn diagram showing the numbers of conserved and strain-specific coding sequences (CDSs). 4,382 core genes shared across all strains, whereas 147 genes were specific to *R. planticola* GODA. (B) COG category-based functional analysis of GODA-specific CDSs. This figure was constructed using Microsoft PowerPoint.

Fig. 3: heat-map of average nucleotide identity values between each genome of *Raoultella planticola* strains and related species. *R. planticola* GODA is closely related to ATCC 33531, FDAARGOS_64, and CHB. This figure was depicted by OrthoANI (https://www.ezbiocloud.net/tools/orthoani).
Fig. 4: genomic comparison of the cps gene cluster in *Raoultella planticola* reveals genetic diversity. Gene clusters are shown in gray. Strain specific *wzx* genes are marked in red color. GT: glycosyltransferase. This figure was constructed using Microsoft PowerPoint.
TABLE II
Phenotypic resistance profile and putative resistance determinant in strain GODA

| Antimicrobial agents (Subclasses) | MIC (µg/mL) | Interpretation | Putative resistance determinant |
|----------------------------------|-------------|----------------|-------------------------------|
| β-lactams                        |             |                |                               |
| Ampicillin/sulbactam              | ≥ 32        | R              | \(blu_{TEM-57}\)               |
| Piperacillin/tazobactam           | ≥ 128       | R              |                               |
| Cefazolin                        | ≥ 64        | R              |                               |
| Cefoperazone/Sulbactam            | ≥ 64        | R              |                               |
| Ceftazidime                      | ≥ 64        | R              |                               |
| Ceftriaxone                      | ≥ 64        | R              |                               |
| Cefepime                         | ≥ 64        | R              |                               |
| Imipenem                         | 4           | R              |                               |
| Ertapenem                        | 2           | R              |                               |
| Aminoglycosides                  |             |                |                               |
| Gentamicin                       | 8           | I              |                               |
| Amikacin                         | ≤ 2         | S              |                               |
| Folate pathway inhibitors        |             |                |                               |
| Trimeproprim/ sulfamethoxazole    | ≥ 320       | R              | \(sul3\)                       |
| Fluoroquinolone                  |             |                |                               |
| Ciprofloxacin                    | 1           | S              | \(mdtK\)                       |

MIC: minimal inhibitory concentration.

Data availability - This genome project has been deposited at the NCBI/GenBank (BioProject PRJNA375797), and includes the raw read data, assembly, and annotation. The assembly is available under accession CP019899; the version described in this paper is version CP019899.

REFERENCES

1. Sekowska A, Raoultella spp. - clinical significance, infections and susceptibility to antibiotics. Folia Microbiol (Praha). 2017; 62(3): 221-7.
2. Bagley ST, Seidler RJ, Brenner DJ. *Klebsiella planticola* sp. nov.: a new species of Enterobacteriaceae found primarily in nonclinical environments. Curr Microbiol. 1981; 6(2): 105-09.
3. Kimura Z, Chung KM, Itoh H, Hiraiishi A, Okabe S. *Raoultella elec trica* sp. nov., isolated from anodic biofilms of a glucose-fed microbial fuel cell. Int J Syst Evol Microbiol. 2014; 64(Pt 4): 1384-8.
4. Drancourt M, Bollet C, Carta A, Rousselier P. Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int J Syst Evol Microbiol. 2001; 51(Pt 3): 925-32.
5. Izard D, Ferragut C, Gavini F, Kersters K, De Ley J, Leclere H. *Klebsiella terrigena*, a new species from soil and water. Int J Syst Evol Microbiol. 1981; 31(2): 116-27.
6. Salmaggi C, Ancona F, Olivetti J, Pagliula G, Ramirez GA. *Raoultella planticola*-associated cholangitis and sepsis: a case report and literature review. QJM. 2014; 107(11): 911-3.
7. Chun S, Yun JW, Huh HJ, Lee NY. Low virulence? Clinical characteristics of *Raoultella planticola* bacteremia. Infection. 2014; 42(5): 899-904.
8. Boattini M, Almeida A, Cardoso C, Cruz CS, Machado C, Vesza Z, et al. Infections on the rise: *Raoultella* spp., clinical and microbiological findings from a retrospective study, 2010-2014. Infect Dis (Lond). 2016; 48(1): 87-91.
9. Momeni SS, Tomlin N, Ruby JD. Isolation of *Raoultella planticola* from refillable antimicrobial liquid soap dispensers in a dental setting. J Am Dent Assoc. 2015; 146(4): 241-5.
10. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016; 66(2): 1112-1117.
11. Rafiq Z, Sam N, Vaidyanathan R. Whole genome sequence of *Klebsiella pneumoniae* U25, a hypermucoviscous, multidrug resistant, biofilm producing isolate from India. Mem Inst Oswaldo Cruz. 2016; 111(2): 857-9.
12. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016; 66(2): 1100-03.
13. Podschan R, Fischer A, Ullman U. Expression of putative virulence factors by clinical isolates of *Klebsiella planticola*. J Med Microbiol. 2000; 49(2): 115-9.
14. Pan YJ, Lin TL, Chen CT, Chen YJ, Hsieh PF, Hsu CR, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. Sci Rep. 2015; 5: 15573.
1. K antigens: relationship to the colanic acid biosynthesis locus and the cps genes from *Klebsiella pneumoniae*. J Bacteriol. 1999; 181(7): 2307-13.

16. Shu HY, Fung CP, Liu YM, Wu KM, Chen YT, Li LH, et al. Genetic diversity of capsular polysaccharide biosynthesis in *Klebsiella pneumoniae* clinical isolates. Microbiology. 2009; 155(Pt 12): 4170-83.

17. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013; 57(7): 3348-57.

18. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012; 40(D1): D115-D22.

19. Hua W, Bao-guang L, Yu-shan P. Prokaryotio expression of TEM-57 typeβ-lactamase gene produced by fowl *Escherichia coli* and its related characteristics. Chin J Vet Sci. 2013; 33(1): 32-7.

20. Sun J, Deng Z, Yan A. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. Biochem Biophys Res Commun. 2014; 453(2): 254-67.

21. Li XZ, Nikaido H. Efflux-mediated drug resistance in bacteria: an update. Drugs. 2009; 69(12): 1555-623.