bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional and Comparative Genomics of Niche-Specific Adapted Actinomycetes *Kocuria rhizophila* Strain D2 Isolated from Healthy Human Gut

Running title: Functional and Comparative Genomes of *Kocuria rhizophila*

Vikas C. Ghattargi *†, Yogesh S. Nimonkar*, Kamala Sape*, Om Prakash*, Mangesh V. Suryavanshi‡, Yogesh S. Shouche*, Bharati S. Meti†§, Shrikant P. Pawar*§

*National Centre for Microbial Resource (NCMR), National Centre for Cell Science (NCCS), Pune- 411021, Maharashtra, India

†Department of Biotechnology, Basaveshwar Engineering College, Bagalkot- 587102, Karnataka, India

‡ Sikkim Central University, Sikkim, India.

§Corresponding Author(s)

Dr. Bharati S. Meti, Department of Bio-technology, Basaveshwar Engineering College, Bagalkot- 587102, Karnataka, India. E mail: bharatismeti@gmail.com.

Dr Shrikant S. Pawar, National Centre for Microbial Resource (NCMR), National Centre for Cell Science (NCCS), Pune- 411021, Maharashtra, India. Email: shrikantpawar@nccs.res.in

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Incidences of infection and occurrence of *Kocuria rhizophila* in human gut are prominent but certainly no reports on the species ability to withstand human gastrointestinal dynamics. *Kocuria rhizophila* strain D2 isolated from healthy human gut was comprehensively characterized. The functional analysis revealed the ability to produce various gastric enzymes and sensitive to major clinical antibiotics. It also exhibited tolerance to acidic pH and bile salts. Strain D2 displayed bile-salt hydrolytic (BSH) activity, strong cell surface traits such as hydrophobicity, auto-aggregation capacity and adherence to human HT-29 cell line. Prominently, it showed no hemolytic activity and was susceptible to the human serum. Exploration of the genome led to the discovery of the genes for the above said properties and has ability to produce various essential amino acids and vitamins. Further, comparative genomics have identified core, accessory and unique genetic features. The core genome has given insights into the phylogeny while the accessory and unique genes has led to the identification of niche specific genes. Bacteriophage, virulence factors and biofilm formation genes were absent with this species. Housing CRISPR and antibiotic resistance gene was strain specific. The integrated approach of functional, genomic and comparative analysis denotes the niche specific adaption to gut dynamics of strain D2. Moreover the study has comprehensively characterized genome sequence of each strain to know the genetic difference and intern recognize the effects of on phenotype and functionality complexity. The evolutionary relationship among strains along and adaptation strategies has been included in this study.

Keywords: commensal, Indian, Gluten, comparative genome analysis, *In-silico* analysis

Significance: Reports of *Kocuria rhizophila* isolation from various sources have been reported but the few disease outbreaks in humans and fishes have been prominent, but no supportive

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evidence about the survival ability of *Kocuria spp.* within human GIT. Here, we report the gut adaption potential of *K. rhizophila* strain D2 by functional and genomic analysis. Further; comparative genomics reveals this adaption to be strain specific (Gluten degradation). Genetic difference, evolutionary relationship and adaptation strategies have been including in this study.

INTRODUCTION

The genus Kocuria (formerly Micrococcus) was named after a Slovakian microbiologist: Miroslav Kocur, and belongs to the class *Actinobacteria* (1). It is Gram-positive cocci, catalase positive, non-hemolytic, non-endospore-forming, non-motile and can grow at different oxygen levels (aerobic, facultative and anaerobic) (1, 2). Presence of galactosamine and glucosamine (amino sugars) as the main component of the cell wall, differ them from other members of the class *Actinobacteria* (2). This genus is normally inhabitants of dust, soil, water and food and in humans colonizes skin, mucosa, oropharynx and gastrointestinal tract (GIT) (3–5).

K. rhizophila was isolated from the rhizoplane of the narrow-leaved cattail (*Typha angustifolia*). in 1999 (6). The strain ATCC 9341 was reclassified as *K. rhizophila* from *Micrococcus luteus* (7). It has been isolated from viz. cheese (8), chicken meat (9) and also healthy human GIT (1) across the globe, thus suggesting its wide adaption potential. Moreover, it has been important in industrial applications for antimicrobial susceptibility testing as standard quality control strain (1–3).

Currently, *K. rhizophila* is gaining importance as emerging pathogen in immune-compromised and metabolically disordered individuals (10–14). In particular, their affinity to plastic materials and devices such as a catheter, causing chronic recurrent bacteremia and thus causing mortality (10–14). Therefore, one should not underestimate the significance of such microorganisms when

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

isolated from clinical samples and particularly from Gastrointestinal tract (GIT), blood and medical implant surfaces. Currently, there are two complete genomes (*K. rhizophila* DC2201 and FDAARGOS_302) and five draft genome (P7-4, TPW45, 14ASP, RF and G2) sequences of *K. rhizophila* available publicly at NCBI (National Center for Biotechnology Information). Based on genome sequence it displays a wide range of activity viz. tolerance to various organic compounds and sturdy amino-acid and carbohydrate metabolism.

Recent techniques of 16S rRNA amplicon and metagenomics sequencing have vividly expanded the known diversity of the human gut microbiome (15–18) but the first approach used to study the gut microbiota employed microbial culture (1). Recent studies with culturomics approach have provided actual insights into the type of species present in the human gut (18–20). Using above-said method (culturomics), we could isolate more than 120 different strains from 18 different genera were isolated; using 35 culture media and different growth conditions.

Here, we present the work carried out on *K. rhizophila* strain D2 isolated from the healthy human gut as there is no supportive evidence about the survival ability of *Kocuria spp.* within human GIT. Thus we found the importance to study strain D2 for its adaption, pathogenicity, commensal or beneficial nature. The work described here utilizes *in-vitro,* genomes and comparative genomics approach to identify the potential of *K. rhizophila* strain D2.

RESULTS

Identification

The 16S rRNA gene of strain D2 showed 99.92% similarity to *Kocuria rhizophila* type strain DSM 11926. The phylogenetic tree was constructed using the Neighbour-Joining method with closely related taxa (Fig 1).

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Characterization of *K. rhizophila* D2

Exoenzymes, Carbohydrate utilization, Antibiogram and Plasmid determination

The strain D2 showed various gastric enzymes viz. lipase, urease, phosphatase, protease, catalase and amylase. The activity was also found for nitrate reductase activities (Table 1). The strain was also able to utilize most of the sugars viz. inulin, lactose, sucrose, fructose, maltose, galactose, dextrose, raffinose, trehalose and melibiose (Table 2). The strain had intermediate resistance to levofloxacin, ciprofloxacin and gentamicin while sensitive to other 21 of the 24 tested antibiotics (Table 3). The plasmid was not associated with the strain.

Bile, Acid and hydrogen peroxide tolerance

Strain D2 was able to grow in the various bile concentration from 0.1- 1 % (w/v), of these it showed 60% survivability at 0.4% bile salt after 24hrs. For the acid tolerance assay strain D2 showed 58.8% survivability at pH 2.5 for 3h. The time depicts the time taken by the food in the stomach. Further, strain D2 was able to tolerate hydrogen peroxide for 4.5 hours.

Adhesion, Exopolysaccharide production and Bile salt hydrolytic (BSH) activity

Auto-aggregation capacity of strain D2 was 29%, while the cell surface hydrophobicity to hydrocarbons viz. toluene, xylene and hexane was 31%, 28% and 22% respectively. The percent cell adhesion assay for human HT-29 cell line was found to be 30%. An additional microscopic observation revealed low adhesion across the quadrants (Fig 2). The strain D2 was able to hydrolyze bile, a zone of clearance (9 mm) was seen when grown on medium with bile salts and this ability was further confirmed by the Ninhydrin method [4 ±0.2 mg/ cell pellet (mg)]. The strain D2 could not produce exopolysaccharide.

Resistance to Simulated GI Conditions and Gluten Degradation

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The survival rate of strain D2 after 3 h exposure to simulated gastric juice was $78x10^8$ cells, which has reduced from $90x10^8$, the zero hour reading. Further, these cells were washed in PBS and added to the simulated intestinal juice. Only $59x10^8$ was alive after three hours. It could also utilize gluten as a sole source of nitrogen, and thus form a clear zone (11±0.1mm) around the colony grown.

Pathogenicity testing

Pathogenicity testing was included in the investigation. Hemolytic activity, serum resistance and biofilm formation have generally been used. The strain D2 was susceptible to human serum with 0.04% survival, not a biofilm producer and exhibited alpha hemolytic activity.

Genome Features of *K. rhizophila* D2

More than 4.9 million good quality paired-end reads were obtained, with an approximate 110x sequencing coverage. The nearly complete genome of *Kocuria rhizophila* D2 consisted of 2,313,294 bp (2.3 Mb) with an average G+C content of 71.0 %. Genome consisted of 2,253 genes and 2,218 ORFs were identified. Within the genome, 46 structural tRNAs, and 3 rRNAs could be predicted.

Genome-based metabolic capabilities of *K. rhizophila* D2

We screened the genome sequence of the strain D2 for various metabolic capabilities occurring within, helping to understand the cellular processes.

Amino Acid Synthesis* and *Proteolytic System

Analysis of genome reports the ability to synthesize amino acids such as serine, cysteine, and aspartate. From these amino acids, seven other could be generated. Complete pathways for essential amino acids such as valine and leucine could be constructed from the genome sequence. Only D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4) and three general peptidase genes

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

could be identified in strain D2. Membrane proteins belonging to (Zinc) metalloendopeptidases was detected. Two copies of *Clp* protease (serine peptidases) were identified. The presence of aminopeptidases (E.C. 3.4.11.1 and E.C 3.4.11.2) and cytosol aminopeptidase *Pep*A were implied in gluten metabolism, found during the gluten utilization experiment.

Carbohydrate Metabolism

The genome of strain D2 encodes a large diversity of genes related to carbohydrate metabolism. Annotation from PATRIC and PGAP has shown the genes present in pyruvate metabolism I & II, glycolysis and gluconeogenesis, TCA cycle, pentose phosphate pathway and serine-glyoxylate cycle. The genes present in strain D2 helps to utilize various sugars viz. lactose, maltose, inositol, sorbitol, mannitol, fructose, adonitol, dextrose, arabitol, galactose, raffinose, rhamnose, trehalose, cellobiose, melibiose, sucrose, L-arabinose, esculin, d-arabinose, citrate, malonate and sorbose. The genomic analysis further countersigns the presence of these genes and their transporters within the genome.

Allied Metabolism

We could identify the entire components of thioredoxin system along with FMN reductase (EC 1.5.1.29) responsible for the sulfur assimilation into the cell. Enzymes exopolyphosphatase (EC 3.6.1.11) and polyphosphate kinase 2 (EC 2.7.4.1) catalyzes the hydrolysis of inorganic polyphosphate and formation of polyphosphate from ATP respectively. The genome also had seven genes responsible for ammonia assimilation into its cells, which chiefly includes glutamine synthesis genes. The genome also has pathways for the production of vitamins: menaquinone (vitamin K2), phylloquinone (vitamin K1), thiamine (B1) riboflavin (B2) and folate (B9). Also, complete pathways for carotenoids and primary bile salts could be identified.

Stress Response

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The genome of *Kocuria rhizophila* D2 encodes various stress-related proteins, involving various proteases involved in the stress response. Highly conserved class I heat-shock genes (GroES and GroES operons) and F_1F_0-ATPase system for maintaining the integrity of cellular proteins under stress conditions were identified. The genome harbours three copies of universal stress proteins, non-specific DNA-binding protein and genes related to oxidative and osmotic stress were found.

Comparative genomics

General genomic features

The average genome sizes of *K. rhizophila* strains were approximately of 2.74Mb and average GC content is 70.81 %. Comparison between GC content, genome sizes, number of genes and coding DNA sequence (CDS), we could not obtain any significant differences ($P \leq 0.05$, Kruskal–Wallis statistical test). The average number of annotated protein-encoding genes is 2345. Analysis of RAST suggests the abundance of amino acids and its derivatives. Table 4 provided the general genome features for the strains.

Comparisons of D2 with other K. rhizophila strains

The availability of various *K. rhizophila* genomes has helped to define a core, accessory, unique genome features. The comparison of strain D2 with other strains revealed 888 (40.11%) core genes, 1243 (56.14 %) accessory and 83(3.74%) unique genes. The core and pan graph for a number of shared gene families to the number of strains is plotted as shown in Fig 3. The size of the core-genome gradually stabilized while the pan-genome size grew continuously, by the addition of other strains indicating an open pan-genome.

Pan-Genome Analysis

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The analysis of pan, core and accessory genome revealed the presence of 888 core, 10882 accessory genes. The number of strain-specific genes for strain 14ASP is 453, strain D2 are 83, DC2201 are 21, FDAARGOS_302 are 35, G2 are 514, P7-4 are 276, RF are 160 and for strain TPW45 are 203 (Fig 4). orthoMCL analysis of **core genes** leads to the identification of copy number within the genomes. 550 genes were present in a single copy and 338 genes were present in multiple copies within eight strains. Functional analysis of the core genes by Cluster of Orthologous Genes (COG) showed the distributed in a varied range of functional categories. These included genes related to cell growth, DNA replication, transcription, translation, and also general transporters. The analysis also revealed presence of carbohydrate, amino acid metabolism, stress response and secondary metabolism. Categories of representing growth, replication, transcription, translation, transporters comprised of 45.12% of the core genes. The core (888) genes were used to construct a phylogenetic tree for all the strains under study, where *K. flava* was used as an out-group. Phylogenetic reconstruction by using (ML) Maximum likelihood method separated the study 8 strains in 2 clusters with bootstrap more than 70 (Fig 5) and the same observation was made when repeated with the pan-genome (data not shown). Clustering based on the source of isolation could not be observed.

Functional analysis of the **accessory genes** shows the limited distribution in COG categories as opposed to core gene annotation. We could identify genes present in secondary metabolism, transport and adaption only. The functional annotation has shown the presence of a large percentage (52.19%) of genes was assigned to an uncharacterized group. The analysis also identified a number of **unique genes** associated with the various strains. The strain 14ASP had the most unique number of genes while the strain DC221 has the least number of unique genes. The annotation of these has helped to identify the role of these genes within them.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We could identify proteins involved in targeting and insertion of nascent membrane proteins into the cytoplasmic membrane, and SRP receptor for *Fts*Y and *Rec*O family involved in involved in DNA repair in strain FDAARGOS_302. In strain TPW45 arginine biosynthesis, bifunctional protein *Arg*J, O-succinylbenzoate synthase, N-acetyl-gamma-glutamyl-phosphate reductase was found. In strain D2 we could find Isopentenyl-diphosphate Delta-isomerase and general transmembrane protein while strain P7_4 had formimidoyl glutamase, single-stranded DNA-binding protein and NADH-quinone oxidoreductase subunit K. In strain 14ASP alanine racemase, cysteine--tRNA ligase, Holliday junction ATP-dependent DNA helicase *Ruv*B, crossover junction endo-de-oxyribonuclease *Ruv*C, lipoprotein signal peptidase; peptide methionine sulfoxide reductase *Msr*A and sec-independent protein translocase protein *Tat*A. While strain G2 had formamidopyrimidine-DNA glycosylase; ATP-dependent dethiobiotin synthetase *Bio*D; ribosome hibernation promoting factor; thiamine-phosphate synthase and strain RF had urease accessory protein *Ure*D while strain RF unique genes were associated with hypothetical proteins. A large portion (97.82%) of these unique genes was assigned to hypothetical proteins while only 2.17% could be assigned to some functions.

Mobile genetic elements (MGE)

A number of MGEs have been described in *K. rhizophila* including transposons, plasmids, and bacteriophage. Based on the screening performed the **IS element** viz. IS481, IS5, TN3 was only present in all other strains except in 14ASP (Table 5). We could identify IS21 in TPW45 and IS21 in P7-4 alone. We could identify maximum of 35 copies of intentional sequences in strain 14ASP and minimum of 7 in strain D2. Further, no **prophage** could be identified in all the eight *K. rhizophila* genomes while Clustered Regularly Interspaced Short Palindromic Repeat (**CRISPR**) were present in 3 strains: 14ASP, TPW45, RF.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomic Islands are distinct DNA fragments associated with mobility and we could identify maximum of 22 GIs in G2. These genomic islands comprised of a minimum of 3% to a maximum of 12.8% of genome size in the strain considered under study. These strain (DC220, FDAARGOS 302, P7-4) had an equal number of genomic islands i.e. 11. The strain D2 has the least number of islands and comprises 3.4% of the total genome. We could identify many important genes associated with a cellular function in these regions. Some of these in strain DC220 are ethanolamine permease and glutamine synthetase; dethiobiotin synthetase and fusaric acid resistance protein in strain FDAARGOS 302; acyl-CoA synthesis genes within P7-4; bleomycin resistance family protein in TPW45; glutaminase synthesis gene in 14ASP; thioredoxin in RF; short-chain dehydrogenase in G2; alkylmercury lyase (EC 4.99.1.2) in D2. Further, comparison of the genomic islands, we could not identify any common GI but could find all the IS elements, CRISPR cas genes within these regions. The genome ATLAS plot shows these differences between the strains (Fig 6).

Antibiotic resistance, Virulence determinants and Survival in GIT

K. rhizophila's report on resistance to antibiotics is very little known. Thus in this study, we screened genomes of *K. rhizophila* against Comprehensive Antibiotic Resistance Database (CARD) for antibiotic resistance genes. Only strain RF had a single copy of beta-lactamase gene and no antibiotic resistance from other seven isolates were identified. No virulence factors could be found in any of the genomes. We could also identify genes *lytr*, *rrp1* for acid resistance, *clp* for bile resistance and *copA* gene for competitiveness have been identified; thus helping it to survive in the gastrointestinal tract within strain P7-4 and D2 only.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION

The isolate *K. rhizophila* D2 was comprehensively characterized for its ability to colonize in the gut and, if present antibiotic resistance and virulence factors. In, the present study we use a combination of *in-vitro* and *in-silico* approaches to identify this potential of bacterial strain D2. A successful colonization in GIT by bacteria can happen if they have the ability to tolerate low pH, bile salts, oxidative stress and moreover survive in the obligate anaerobic environment (21, 22).

Principally, we tested the acid and bile tolerance and found that strain D2 was able to tolerate a low pH of 2.5 and 0.4% (w/v) bile salts with 60% and 58.8% survivability. Along with experimental evidence, the genes were identified involved in acid tolerance (*lyt*, *clp*) within the genome (23). We also identified genes encoding for entire primary bile salts production pathway responsible for bile salts hydrolysis activity and in-turn the tolerance. Thus indicating important characters for a bacterium to stay alive and become part of the natural GIT microbial community.

Next, we examined the aggregation and adhesion properties of strain that play an important attribute for long-term colonization in the human GIT (24). The strain D2 showed low autoaggregation capacity and adhesion. The isolate was also to adhere to human HT-29 cells which were evident from the cells surface hydrophobicity. Also, the genome of strain D2 has *copA* gene which helps in competitiveness with other bacteria (23). The resistance to hydrogen peroxide is imparted by iron-binding ferritin-like antioxidant protein and superoxide dismutase (EC 1.15.1.1) (25) found in the genome, this activity was also shown in the experiments. Overall these ability suggest its potential to thrive in the GIT conditions. Further, we could not find genes producing exopolysaccharides (EPS) and *in-vitro* results for production of EPS by strain D2

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

confirm its inability to produce EPS. Generally these EPS are meant to provide additional benefit as a means of protection from various stress conditions present in the human GIT (26, 27).

Genes for carbohydrate (CHO) metabolism revealed a diverse range of the gene encoding for utilizing numerous carbon sources that can be used for energy and growth. We found important genes found in CHO metabolism along with their respective transport systems. Further, the API stress assay reflected the functionally of the genome. The integrated approach of genomic and functional features of the strain together provides a comprehensive mechanism of carbohydrate utilization. The *in-silico* analysis also revealed the presence of 'Opp proteins' proteolytic system that help in breaking of high molecular weight proteins into smaller proteins, thus converting to absorbable forms for the human body (28–30). The utilization of gluten as a sole source of nitrogen was seen from the aminopeptidases genes present in the genome, which is a unique property for the strain D2.

The capacity to absorb sulphur, phosphate and nitrogen was found to be related with the genome by strain D2, helping them to carry out its own cellular process. The strain D2 has potential to synthesize menaquinone (vitamin K2), phylloquinone (vitamin K1), thiamine (B1) riboflavin (B2), folate (B9). These vitamins are need to be supplied exogenously and cannot be synthesized by the human cells thus essential (31, 32). Moreover, it can produce in eleven amino acids which serve as precursors for the synthesis of short-chain fatty acids (33, 34). The genome has genes for acetyl-, butyryl- and proponyl- CoA dehydrogenase but the last enzymes that convert these into acetate, buterate and propionate were absent. This is evident as this bacterium belongs to actinomycetes and acetyl-, butyryl- and proponyl- CoA dehydrogenase are further taken for another process (35, 36).

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

One concern regarding the strains of *K. rhizophila* is that this species has been known for its pathogenic infections. Therefore, it is of high importance to test for pathogenicity. Strain D2 showed non hemolytic activity and was sensitive to human serum and showed marginally sensitivity to some other antibiotic such as ciprofloxacin, levofloxacin, and gentamicin Thus, the strain D2 is a non-hemolytic, sensitive to major antibiotics tested and as well we could not identify the hemolytic genes in genome analysis.

Genome comparison did not reveal any significant differences ($P \leq 0.05$) between the strains with reference to their GC content, genome size, an average number of genes and coding DNA sequence (CDS). The pan-genome size grew steadily with the addition of strains and the core genome stabilized, thus indicating an open pan-genome for the strains under study (Fig 4). The pan-genome analysis revealed 888 (6.57%) as core genes, 10882 (80.51%) accessory and 1745 (12.91%) as unique genes. The less number of genes in core, unique category and a large number of genes in accessory suggests the genomic fluidity of the genomes (37, 38). Further, the phylogenetic tree based on the core genome SNP based phylogeny separated 8 strains in 2 distinct clusters (bootstrap >70) (Fig 5) and no clustering based on the source of isolation.

Most of the strains under study did not harbour any antibiotic resistance gene except strain RF possessing beta-lactamase, indicating its multi-drug resistance to antibiotics such as penicillins, cephalosporins, cephamycins, and carbapenems (39, 40). No virulence genes could be identified in any strains and genes responsible for survival in GIT can be only found in strains P7-4 and D2 as these are the only gut isolates.

Insertion sequences (ISs) and bacteriophages contribute actively to bacterial evolution by integrating and exciting from the genome. In certain conditions, they provide new genetic

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

properties such as virulence factors and antibiotic resistance (41–45) . In *K. rhizophila* no such observations for IS elements with respect to virulence and antibiotic determinates could be done, also the bacteriophage did not harbour important genes related to the bacterial cellular functions. Genomic Islands are DNA fragments which usually are associated with mobility and differ between closely related strains (46, 47). These genomic islands compromised a minimum 3.2% to a maximum of 12% of the genomes in *Kocuria rhizophila* and most of these genes were assigned to hypothetical proteins. Clustered Regularly Interspaced Short Palindromic Repeat (**CRISPR**) was present in three environment strains: TPW45, 14ASP and RF. This has been attributed to the higher frequency of phage attacks present in the environments (48, 49).

In conclusion, the trio approach of *in-vitro* characterization, genome mining and comparative genomics of strain D2 have helped in the perceptive knowledge of genes responsible for surviving in the gut. Moreover, the pan-genome analysis has shown the niche-specific genes responsible for adaption and the pan-genome is open constructed on the bases of eight genomes. The unique genes present the strains D2 ability to stay within the gut and might have the potential to act as potential probiotic, as the strain produces various essential amino acids and vitamins benefiting humans. The important factor of these sequenced genomes implies the absence of virulence factors and biofilm formation ability and antibiotic resistance genes (except strain RF).

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Isolation and Preservation

The approval from Institutional Ethics Committee (IEC) was obtained. Three self-declared healthy volunteers were selected with consent prior to collection of the samples. We immediately transported the collected faecal samples to the lab at 4^{o}C and processed for isolation of faecal bacteria within 6 h. Faecal samples (1g) were transferred to 9 ml of sterile saline (0.85% sodium chloride, Sigma) and mixed well (50). The serial dilutions were subsequently prepared in sterile saline, and appropriate dilutions of the samples plated on Nutrient Agar (HiMedia, Mumbai, India). Plates were incubated at 37^{o}C for 48 h under aerobic condition. Glycerol [20% (v/v)] stocks were prepared to preserve the isolated pure cultures and froze at -80^{o}C (50).

Identification

Genomic DNA of the pure culture was extracted and quantified by using Qiagen Blood & Cell Culture (Qiagen, USA) and Nanodrop ND1000 (Thermo Scientific, USA) respectively. We amplified 16S rRNA gene by using universal primers, 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R (5′-ACG GCT ACC TTG TTA CGA CTT-3′) as described in the earlier study (50). The amplified PCR products were purified using polyethylene glycol (PEG)–NaCl precipitation (50) followed by sequencing in ABI 3730xl DNA analyzer, with the help Big Dye terminator kit (Applied Biosystems, Inc., Foster City, CA). The sequence obtained was assembled using DNASTARPro, version 10 and taxonomic identity were checked by using the EZ-Biocloud server (50, 51).

Characterization of K. rhizophila D2

We used Dodeca Universal I & II kit (Hi-media, India) disc diffusion method, to check antibiotic susceptibility. We tested the isolate for nine exoenzymes by standard microbiological methods

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

viz. Phosphatase by Pikovasky's agar base (52), lipase by tributyrin agar base (53), urease by urease agar base (54), protease by skimmed milk agar (55), gelatinase by gelatin medium (56), cellulase by CMC agar (57), amylase by starch iodine test (58), Nitrate reductase by colour change method (59) and catalase by effervescence of 6% H_2O_2(60). Carbohydrate utilization was conducted by HiCarbohydrateTM Kit (Sigma, India) consisting of thirty-four carbohydrates with respect to manufactures instructions. We also checked for the presence of any plasmid by Qiagen Plasmid Mini Kit (Qiagen, USA). The bile and acid tolerance assay (61–64) autoaggregation(65), cell surface hydrophobicity (66), adhesion to human HT-29 cell line (67, 68),bile-salt hydrolytic (BSH) activity (69), resistance to hydrogen peroxide (70–72) exopolysaccharide production (73), hemolytic activity (74–76), resistance to simulated GIT conditions (77–79), serum resistance (80), resistance to Simulated GI Conditions (77, 81, 82) and gluten degradation (83) was carried out as stated.

Statistical analysis

All the experiments were done in triplicates and the mean values and standard deviation was obtained and Duncan's Multiple Range Test (SPSS Ver. 10.0) was used for comparisons.

Genome Sequencing and Assembly

We extracted genomic DNA as per the manufacturer's protocol (QIAamp genomic DNA kit, Germany). The high-quality DNA was sequenced using Illumina MiSeq platform (2x300 paired-end libraries). PATRIC was used for de-novo assembly of quality-filtered reads (84, 85).

Bioinformatics Analyses

The draft genome sequence was annotated using RAST and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (86). Protein coding genes, tRNA and rRNA genes from the

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

374 genomes were predicted using Glimmer version 3.02 (87), tRNA_scan-SE (88) and RNAmmer
375 (89) respectively. We also use COG database to analyze protein-coding genes (90) and Pfam
376 domains were predicted using NCBI Batch CD-Search Tool (91). Presence of CRISPR repeats
377 was predicted using the CRISPRFinder tool (92). We obtained open reading frames (ORFs) by
378 using the ORF finder tool (https://www.ncbi.nlm.nih.gov/orffinder/). Prophage sequences were
379 predicted and annotated using PHASTER (93). Bacterial insertion elements (ISs) were identified
380 by ISfinder(94). Horizontal gene transfer was detected by genomic island tool: Islandviewer(95).
381 Gene clusters of any bioactive compounds were identified by antiSMASH: antibiotics and
382 Secondary Metabolite Analysis Shell (96). PlasmidFinder was used to search for plasmids within
383 the genome (97). We used PATRIC to predict the metabolic pathways from the genome (98).
384 Comparative genome analysis of ten whole genome sequences of *kocuria rhizophila* was done
385 by an ultra-fast bacterial pan-genome analysis pipeline (BPGA) (99). A blast atlas was generated
386 with the help of GVIEW Server (https://server.gview.ca/) (100).

387 **Accession number(s)**

388 We have deposited this whole genome shotgun project at GenBank under the accession
389 PNRK00000000. The version described in this paper is version PNRK00000000.GenBank
390 accession number for the partial 16S rRNA nucleotide sequence is MH005095.

391

392 **Funding:** This work was supported by the Department of Biotechnology, Government of India,
393 under Grant National Centre for Microbial Resources (NCMR) (BT/Coord.II/01/03/2016).
394 **Acknowledgements:** Additional financial support was provided by the Department of
395 Biotechnology, Basaveshwar Engineering College, Bagalkot, under TEQUIP Grant to Bharati S.
396 Meti.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Savini V, Catavitello C, Masciarelli G, Astolfi D, Balbinot A, Bianco A, Febbo F, D'Amario C, D'Antonio D. 2010. Drug sensitivity and clinical impact of members of the genus Kocuria. J Med Microbiol 59:1395–1402.

2. Zeng M, Pu D, Mo X, Zhu C, Gong S, Xu Y, Lin G, Wu B, He S, Jiao X, Wang X, Zhu Q, Altmeyer R. 2013. Children of rural-to-urban migrant workers in China are at a higher risk of contracting severe hand, foot and mouth disease and EV71 infection: A hospital-based study. Emerg Microbes Infect 2:0.

3. Becker K, Rutsch F, Uek??tter A, Kipp F, K??nig J, Marquardt T, Peters G, Von Eiff C. 2008. Kocuria rhizophila adds to the emerging spectrum of micrococcal species involved in human infections. J Clin Microbiol 46:3537–3539.

4. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin E V., La Scola B, Raoult D. 2012. Microbial culturomics: Paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185–1193.

5. Pękala A, Paździor E, Antychowicz J, Bernad A, Głowacka H, Więcek B, Niemczuk W. 2018. Kocuria rhizophila and Micrococcus luteus as emerging opportunist pathogens in brown trout (Salmo trutta Linnaeus, 1758) and rainbow trout (Oncorhynchus mykiss Walbaum, 1792). Aquaculture 486:285–289.

6. Kovacs G, Burghardt J, Pradella S, Schumann P, Stackebrandt E, Marialigeti K. 1999. Kocuria palustris sp. nov. and Kocuria rhizophila sp. nov., isolated from the rhizoplane of

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

418 the narrow-leaved cattail (Typha angustifolia). Int J Syst Bacteriol 49:167–173.

419 7. Tang JS, Gillevet PM. 2003. Reclassification of ATCC 9341 from Micrococcus luteus to
420 Kocuria rhizophila. Int J Syst Evol Microbiol 53:995–997.

421 8. Anang DM, Rusul G, Radu S, Bakar J, Beuchat LR. 2006. Inhibitory effect of oxalic acid
422 on bacterial spoilage of raw chilled chicken. J Food Prot 69:1913–1919.

423 9. El-Baradei G, Delacroix-Buchet A, Ogier JC. 2007. Biodiversity of bacterial ecosystems
424 in traditional Egyptian Domiati cheese. Appl Environ Microbiol 73:1248–1255.

425 10. von Eiff C, Kuhn N, Herrman M, Weber S, Peters G. 1996. Micrococcus luteusas a cause
426 of recurrent bacteremia. Pediatr J Infect Dis 15:7–10.

427 11. Oudiz RJ, Widlitz A, Beckmann XJ, Camanga D, Alfie J, Brundage BH, Barst RJ. 2004.
428 Micrococcus-associated central venous catheter infection in patients with pulmonary
429 arterial hypertension. Chest 126:90–94.

430 12. Moissenet D, Becker K, Mérens A, Ferroni A, Dubern B, Vu-Thien H. 2012. Persistent
431 bloodstream infection with Kocuria rhizophila related to a damaged central catheter. J
432 Clin Microbiol 50:1495–1498.

433 13. Kogure Y, Nakamura F, Nukina A, Ichikawa M, Kurokawa M, Kamikubo Y. 2014.
434 Catheter-related septic shock by Micrococcus in an autologous hematopoietic stem cell
435 transplantation recipient. Am J Infect Control 42:87.

436 14. Peces R, Gago E, Tejada F, Laures AS, Alvarez-Grande J. 1997. Relapsing bacteraemia
437 due to Micrococcus luteus in a haemodialysis patient with a Perm-Cath catheter. Nephrol

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

438 Dial Transplant 12:2428–2429.

439 15. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy
440 HH, Earl AM, Fitzgerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA,
441 Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J,
442 Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO,
443 Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi
444 HM, Armitage G, Arze CA, Ayyaz T, Baker CC, Begg L, Belachew T, Bhonagiri V,
445 Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam
446 DA, Campbell JL, Canon SR, Cantarel BL, Chain PSG, Chen IMA, Chen L, Chhibba S,
447 Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA,
448 Davidovics NJ, Davis CC, Desantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E,
449 Ding Y, Dooling DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL,
450 Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney
451 LJ, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H,
452 Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A,
453 Griggs A, Gujja S, Kinder Haake S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA,
454 Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson
455 JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB,
456 Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS,
457 Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu
458 Y, Lo CC, Lozupone CA, Dwayne Lunsford R, Madden T, Mahurkar AA, Mannon PJ,
459 Mardis ER, Markowitz VM, Mavromatis K, McCorrison JM, McDonald D, McEwen J,
460 McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ,

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

461 Newsham I, Nusbaum C, Oglaughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM,
462 Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Pop M, Priest ME, Proctor LM, Qin
463 X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-
464 Mueller B, Rogers YH, Ross MC, Russ C, Sanka RK, Sankar P, Fah Sathirapongsasuti J,
465 Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N,
466 Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA,
467 Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG,
468 Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA,
469 Walker J, Wang L, Wang Z, Ward D V., Warren W, Watson MA, Wellington C,
470 Wetterstrand KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C,
471 Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD,
472 Birren BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock
473 GM, Wilson RK, White O. 2012. Structure, function and diversity of the healthy human
474 microbiome. Nature 486:207–214.

475 16. Nih T, Working HMP. 2009. The NIH Human Microbiome Project. Genome Res
476 19:2317–2323.

477 17. Turnbaugh PJ, Ley RE, Hamady M, Fraser-liggett C, Knight R, Gordon JI. 2007. The
478 human microbiome project: exploring the microbial part of ourselves in a changing world.
479 Nature 449:804–810.

480 18. Hugon P, Lagier J-C, Colson P, Bittar F, Raoult D. 2017. Repertoire of human gut
481 microbes. Microb Pathog 106:103–112.

482 19. Greub G. 2012. Culturomics: A new approach to study the human microbiome. Clin

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Microbiol Infect 18:1157–1159.

20. Lagier JC, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, Caputo A, Cadoret F, Traore SI, Seck EH, Dubourg G, Durand G, Mourembou G, Guilhot E, Togo A, Bellali S, Bachar D, Cassir N, Bittar F, Delerce J, Mailhe M, Ricaboni D, Bilen M, Dangui Nieko NPM, Dia Badiane NM, Valles C, Mouelhi D, Diop K, Million M, Musso D, Abrahão J, Azhar EI, Bibi F, Yasir M, Diallo A, Sokhna C, Djossou F, Vitton V, Robert C, Rolain JM, La Scola B, Fournier PE, Levasseur A, Raoult D. 2016. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat Microbiol 1.

21. González-Rodríguez I, Ruiz L, Gueimonde M, Margolles A, Sánchez B. 2013. Factors involved in the colonization and survival of bifidobacteria in the gastrointestinal tract. FEMS Microbiol Lett 340:1–10.

22. Papadimitriou K, Zoumpopoulou G, Foligné B, Alexandraki V, Kazou M, Pot B, Tsakalidou E. 2015. Discovering probiotic microorganisms: In vitro, in vivo, genetic and omics approaches. Front Microbiol 6:1–28.

23. Lebeer S, Vanderleyden J, De Keersmaecker SCJ. 2008. Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev 72:728–764, Table of Contents.

24. Collado MC, Meriluoto J, Salminen S. 2008. Adhesion and aggregation properties of probiotic and pathogen strains. Eur Food Res Technol 226:1065–1073.

25. Fukai T, Ushio-Fukai M. 2011. Superoxide Dismutases: Role in Redox Signaling, Vascular Function, and Diseases. Antioxid Redox Signal 15:1583.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

504 26. Limoli DH, Jones CJ, Wozniak DJ, Cruz S. 2015. Bacterial Extracellular Polysaccharides
505 in Biofilm Formation and Function. Microbiol Spectr 3:1–30.

506 27. van Bueren AL, Saraf A, Martens EC, Dijkhuizena L. 2015. Differential metabolism of
507 exopolysaccharides from probiotic lactobacilli by the human gut symbiont Bacteroides
508 thetaiotaomicron. Appl Environ Microbiol 81:3973–3983.

509 28. Medrano MS, Ding Y, Wang XG, Lu P, Coburn J, Hu LT. 2007. Regulators of expression
510 of the oligopeptide permease A proteins of Borrelia burgdorferi. J Bacteriol 189:2653–
511 2659.

512 29. Hiron A, Borezée-Durant E, Piard JC, Juillard V. 2007. Only one of four oligopeptide
513 transport systems mediates nitrogen nutrition in Staphylococcus aureus. J Bacteriol
514 189:5119–5129.

515 30. Borezee E, Pellegrini E, Berche P. 2000. OppA of Listeria monocytogenes , an
516 Oligopeptide-Binding Protein Required for Bacterial Growth at Low Temperature and
517 Involved in Intracellular Survival OppA of Listeria monocytogenes , an Oligopeptide-
518 Binding Protein Required for Bacterial Growth at Low . Society 68:7069–7077.

519 31. Thomas DR. 2006. Vitamins in aging, health, and longevity. Clin Interv Aging 1:81–91.

520 32. Mora JR, Iwata M, Andrian UH Von. 2008. Vitamin effects on the immune system. Nat
521 Rev Immunol 8:685–698.

522 33. Neis EPJG, Dejong CHC, Rensen SS. 2015. The role of microbial amino acid metabolism
523 in host metabolism. Nutrients 7:2930–46.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

524 34. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. 2013.
525 The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host
526 energy metabolism. J Lipid Res 54:2325–2340.

527 35. Shapiro S (Stuart). 1989. Regulation of secondary metabolism in actinomycetes. CRC
528 Press.

529 36. Rodríguez E, Banchio C, Diacovich L, Bibb MJ, Gramajo H. 2001. Role of an essential
530 acyl coenzyme A carboxylase in the primary and secondary metabolism of Streptomyces
531 coelicolor A3(2). Appl Environ Microbiol 67:4166–76.

532 37. Kislyuk AO, Haegeman B, Bergman NH, Weitz JS. 2011. Genomic fluidity: An
533 integrative view of gene diversity within microbial populations. BMC Genomics 12.

534 38. Ahmed N, Dobrindt U, Hacker J, Hasnain SE. 2008. Genomic fluidity and pathogenic
535 bacteria: Applications in diagnostics, epidemiology and intervention. Nat Rev Microbiol
536 6:387–394.

537 39. Shrestha A, Bajracharya AM, Subedi H, Turha RS, Kafle S, Sharma S, Neupane S,
538 Chaudhary DK. 2017. Multi-drug resistance and extended spectrum beta lactamase
539 producing Gram negative bacteria from chicken meat in Bharatpur Metropolitan, Nepal.
540 BMC Res Notes 10:1–5.

541 40. Chagas TPG, Seki LM, Cury JC, Oliveira JAL, Dávila AMR, Silva DM, Asensi MD.
542 2011. Multiresistance, beta-lactamase-encoding genes and bacterial diversity in hospital
543 wastewater in Rio de Janeiro, Brazil. J Appl Microbiol 111:572–581.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

544 41. Wu Y, Aandahl RZ, Tanaka MM. 2015. Dynamics of bacterial insertion sequences: Can
545 transposition bursts help the elements persist? Theories and models. BMC Evol Biol 15:1–
546 12.

547 42. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: Their genomic
548 impact and diversity. FEMS Microbiol Rev 38:865–891.

549 43. Koskella B, Brockhurst MA. 2014. Bacteria-phage coevolution as a driver of ecological
550 and evolutionary processes in microbial communities. FEMS Microbiol Rev 38:916–931.

551 44. Klimenko AI, Matushkin YG, Kolchanov NA, Lashin SA. 2016. Bacteriophages affect
552 evolution of bacterial communities in spatially distributed habitats: A simulation study.
553 BMC Microbiol 16.

554 45. Tao P, Wu X, Rao V. 2018. Unexpected evolutionary benefit to phages imparted by
555 bacterial CRISPR-Cas9. Sci Adv 4:eaar4134.

556 46. Lacour S, Gaillard M, Van Der Meer JR. 2006. Genomic islands and evolution of
557 catabolic pathways. Prokaryotic Divers Mech Significance Publ Soc Gen Microbiol
558 9780521869:255–274.

559 47. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. 2014. Conjugative and mobilizable
560 genomic islands in bacteria: Evolution and diversity. FEMS Microbiol Rev 38:720–760.

561 48. Haerter JO, Trusina A, Sneppen K. 2011. Targeted Bacterial Immunity Buffers Phage
562 Diversity. J Virol 85:10554–10560.

563 49. 2003. Bacterial Protection Mechanisms against Phage Attack : An Evolutionary

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

564 Perspective.

565 50. Ghattargi VC, Nimonkar YS, Burse SA, Davray D, Kumbhare S V. 2018. Genomic and
566 physiological analyses of an indigenous strain , Enterococcus faecium 17OM39. Funct
567 Integr Genomics.

568 51. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won
569 S, Chun J. 2012. Introducing EzTaxon-e: A prokaryotic 16s rRNA gene sequence
570 database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol
571 62:716–721.

572 52. Shen H, He X, Liu Y, Chen Y, Tang J, Guo T. 2016. A complex inoculant of N2-fixing,
573 P- and K-solubilizing bacteria from a purple soil improves the growth of kiwifruit
574 (Actinidia chinensis) plantlets. Front Microbiol 7:1–12.

575 53. Yusof M, Samad A, Nyonya C, Razak A, Salleh AB, Zin WM, Yunus W, Ampon K, Basri
576 M, Biokimia J, Mikrobiologi D, Kimia 2jabatan, Sains F, Pengajian D, Sekitar A. 1989. A
577 plate assay for primary screening of lipase activity. J Microbiol Methods 9:5156.

578 54. Vuye A, Pijck J. 1973. Urease Activity of Enterobacteriaceae: Which Medium to Choose.
579 Appl Microbiol 26:850–854.

580 55. Savijoki K, Ingmer H, Varmanen P. 2006. Proteolytic systems of lactic acid bacteria. Appl
581 Microbiol Biotechnol 71:394–406.

582 56. Whaley DN, Dowell VR, Wanderlinder LM, Lombard GL. 1982. Gelatin agar medium for
583 detecting gelatinase production by anaerobic bacteria. J Clin Microbiol 16:224–229.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

584 57. Kasana RC, Salwan R, Dhar H, Dutt S, Gulati A. 2008. A rapid and easy method for the
585 detection of microbial cellulases on agar plates using Gram's iodine. Curr Microbiol
586 57:503–507.

587 58. Carrasco M, Villarreal P, Barahona S, Alcaíno J, Cifuentes V, Baeza M. 2016. Screening
588 and characterization of amylase and cellulase activities in psychrotolerant yeasts. BMC
589 Microbiol 1–9.

590 59. Tiso M, Schechter AN. 2015. Nitrate reduction to nitrite, nitric oxide and ammonia by gut
591 bacteria under physiological conditions. PLoS One 10:1–18.

592 60. Whittenbury R. 1964. Hydrogen Peroxide Formation and Catalase Activity in the Lactic
593 Acid Bacteria. J Gen Microbiol 35:13–26.

594 61. Hassanzadazar H, Ehsani A, Mardani K, Hesari J. 2012. Investigation of antibacterial ,
595 acid and bile tolerance properties of lactobacilli isolated from Koozeh cheese. Vet Res
596 Forum 3:181–185.

597 62. Chou LS, Weimer B. 1999. Isolation and characterization of acid- and bile-tolerant
598 isolates from strains of Lactobacillus acidophilus. J Dairy Sci 82:23–31.

599 63. Tokatl M, Gülgör G, Baɬder Elmac S, Arslankoz Işleyen N, Özçelik F. 2015. In Vitro
600 Properties of Potential Probiotic Indigenous Lactic Acid Bacteria Originating from
601 Traditional Pickles. Biomed Res Int 2015.

602 64. Muller JA, Ross RP, Sybesma WFH, Fitzgerald GF, Stanton C. 2011. Modification of the
603 technical properties of lactobacillus johnsonii NCC 533 by supplementing the growth

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

medium with unsaturated fatty acids. Appl Environ Microbiol 77:6889–6898.

65. Schmidt H, Hensel M. 2004. Pathogenicity islands in bacterial pathogenesis. ClinMicrobiolRev 17:14–56.

66. Saran S, Bisht MS, Singh K. 2012. Comparing Adhesion Attributes of two Isolates of Lactobacillus Acidophilus for Assessment of Prebiotics ,. Int J Sci Res Publ 2:2–8.

67. Gyles C, Boerlin P. 2014. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. Vet Pathol 51:328–40.

68. Mishra M, Nimonkar Y, Shouche YS, Kumbhare S, Vaishampayan A, Josef N, Prakash O. 2015. Pantoea intestinalis sp. nov., isolated from the human gut. Int J Syst Evol Microbiol 65:3352–3358.

69. Zanotti I, Turroni F, Piemontese A, Mancabelli L, Milani C, Viappiani A, Prevedini G, Sanchez B, Margolles A, Elviri L, Franco B, van Sinderen D, Ventura M. 2015. Evidence for cholesterol-lowering activity by Bifidobacterium bifidum PRL2010 through gut microbiota modulation. Appl Microbiol Biotechnol 99:6813–6829.

70. Wu D, Sun MZ, Zhang C, Xin Y. 2014. Antioxidant properties of Lactobacillus and its protecting effects to oxidative stress caco-2 cells. J Anim Plant Sci 24:1766–1771.

71. Halliwell B, Clement MV, Long LH. 2000. Hydrogen peroxide in the human body. FEBS Lett 486:10–13.

72. Forman HJ, Bernardo A, Davies KJA. 2016. What is the concentration of hydrogen peroxide in blood and plasma? Arch Biochem Biophys 603:48–53.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

624 73. Abdhul K, Ganesh M, Shanmughapriya S, Kanagavel M, Anbarasu K, Natarajaseenivasan
625 K. 2014. Antioxidant activity of exopolysaccharide from probiotic strain Enterococcus
626 faecium (BDU7) from Ngari. Int J Biol Macromol 70:450–454.

627 74. Semedo T, Santos MA, Martins P, Silva Lopes MF, Figueiredo Marques JJ, Tenreiro R,
628 Barreto Crespo MT. 2003. Comparative study using type strains and clinical and food
629 isolates to examine hemolytic activity and occurrence of the cyl operon in enterococci. J
630 Clin Microbiol 41:2569–2576.

631 75. Semedo T, Almeida Santos M, Martins P, Silva Lopes MF, Figueiredeo Marques, J. J.
632 Teneiro R, et al. 2003. Comparative study using type strains and clinical and food isolated
633 to examine hemolytic activity and ocurrente of the cyl operon in enterococci. J Clin
634 Microbiol 41:2569–2576.

635 76. Ike Y, Hashimoto H, Clewell DONB. 1987. High incidence of hemolysin production by
636 Enterococcus (Streptococcus) faecalis strains associated with human parenteral
637 infections . High Incidence of Hemolysin Production by Enterococcus (Streptococcus)
638 faecalis Strains Associated with Human Parente. J Clin Microbiol 25:1524–1528.

639 77. Vamanu E. 2017. Effect of gastric and small intestinal digestion on lactic acid bacteria
640 activity in a GIS1 simulator. Saudi J Biol Sci 24:1453–1457.

641 78. Izquierdo E, Medina M, Ennahar S, Marchioni E, Sanz Y. 2008. Resistance to simulated
642 gastrointestinal conditions and adhesion to mucus as probiotic criteria for Bifidobacterium
643 longum strains. Curr Microbiol 56:613–618.

644 79. Casarotti SN, Todorov SD, Penna ALB. 2015. Effect of different matrices on probiotic

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

645 resistance to in vitro simulated gastrointestinal conditions. Int J Dairy Technol 68:595–
646 601.

647 80. King LB, Swiatlo E, Swiatlo A, McDaniel LS. 2009. Serum resistance and biofilm
648 formation in clinical isolates of Acinetobacter baumannii. FEMS Immunol Med Microbiol
649 55:414–421.

650 81. Pan X, Wu T, Zhang L, Cai L, Song Z. 2009. Influence of oligosaccharides on the growth
651 and tolerance capacity of lactobacilli to simulated stress environment. Lett Appl Microbiol
652 48:362–367.

653 82. Corcoran BM, Stanton C, Fitzgerald GF, Ross RP. 2005. Survival of Probiotic
654 Lactobacilli in Acidic Environments Is Enhanced.pdf 71:3060–3067.

655 83. Rizzello CG, De Angelis M, Di Cagno R, Camarca A, Silano M, Losito I, De Vincenzi M,
656 De Bari MD, Palmisano F, Maurano F, Gianfrani C, Gobbetti M. 2007. Highly efficient
657 gluten degradation by lactobacilli and fungal proteases during food processing: New
658 perspectives for celiac disease. Appl Environ Microbiol 73:4499–4507.

659 84. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM,
660 Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK,
661 Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M,
662 Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the
663 all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res
664 45:D535–D542.

665 85. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

R, Hix D, Kenyon R, MacHi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42:581–591.

86. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624.

87. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641.

88. Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for inproved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.

89. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108.

90. Tatusov RL, Galperin MY, Natale DA, Koonin E V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–6.

91. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V., Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

687 Res 39:225–229.

688 92. Grissa I, Vergnaud G, Pourcel C. 2008. CRISPRcompar: a website to compare clustered
689 regularly interspaced short palindromic repeats. Nucleic Acids Res 36:52–57.

690 93. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a
691 better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:1–6.

692 94. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference
693 centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-6.

694 95. Juhas M, Van Der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. 2009.
695 Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS
696 Microbiol Rev 33:376–393.

697 96. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber
698 T, Takano E, Breitling R. 2011. AntiSMASH: Rapid identification, annotation and
699 analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal
700 genome sequences. Nucleic Acids Res 39:339–346.

701 97. Carattoli A, Zankari E, García-fernández A, Larsen V, Lund O, Villa L, Aarestrup M,
702 Hasman H. 2014. In Silico Detection and Typing of Plasmids using PlasmidFinder and
703 Plasmid Multilocus Sequence Typing 58:3895–3903.

704 98. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto
705 encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34.

706 99. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S,

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

100. Petkau A, Stuart-Edwards M, Stothard P, van Domselaar G. 2010. Interactive microbial genome visualization with GView. Bioinformatics 26:3125–3126.

FIGURES LEGENDS

Fig 1 Phylogenetic relationship of strain D2 with closely related taxa based on 16S rRNA gene sequences (16S rRNA gene sequence of *Arthrobacter roseus* was used as out-group). The phylogenetic trees were constructed using the Neighbour-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1)

Fig 2 Results for cell adhesion assay performed on Human HT-29 Cell line. Images taken under fluorescence microscope 60X oil immersion where panel (a) is negative control without any bacterial cell added; panel (b) is for the strain D2

Fig 3 Core and pan genome for *K. rhizophila* strains. The number of shared genes is plotted as the function of number of strains (n) added sequentially. 888 copy genes are shared by all 8 genomes. The orange line represents the least-squares fit to the power law function f(x)=a.x^b where a= 2705.9, b= 0.373747. The red line represents the least-squares fit to the exponential decay function f1(x)=c.e^(d.x) where c= 2196.32, d= -0.179618.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 4 Number of Core, Accessory and Unique gene families of *K. rhizophila* genomes. The inner octagon represents the core genome consisting of 888 genes. The outer octagon represents the accessory genome and the number within them represents number of genes, while the triangles represent the unique genes associated with all the strains.

Fig 5 Core Genome Phylogeny. Phylogenetic tree of 8 *K. rhizophila* strains using the Maximum Likelihood method based on the GTR + G substitution model. The tree with the highest log likelihood (-17897.1414) is shown. Evolutionary analyses were conducted in MEGA6. A concatenated tree of 7909 genes was considered in the final dataset.

Fig 6. Blast Atlas of *Kocuria* genomes, with strain FDAARGOS_302 as a reference genome followed by DC2201, G2, TPW45, P7_4, RF and the outermost as D2. The difference between these genomes can be seen by the gaps in the rings.

TABLES

Table 1 Exoenzymes produced by *Kocuria rhizophila* D2

Sl No.	Exoenzyme	Result	Sl No.	Exoenzyme	Result	Sl No.	Exoenzyme	Result
1	Phosphatase	+	4	Cellulase	-	7	Catalase	+
2	Urease	+	5	Protease	+	8	Amylase	+
3	Lipase	+	6	Gelatinase	-	9	Nitrate reductase	+

+ sign indicates enzyme produced and - sign indicates that it does not produce enzymes

Table 2 Various carbohydrates utilised by test strain *Kocuria rhizophila* D2

Sl. No.	Carbohydrates	Strain D2	Sl. No.	Carbohydrates	Strain D2
1	Lactose	Y	18	Inositol	Y
2	Xylose	N	19	Sorbitol	Y
3	Maltose	Y	20	Mannitol	Y
4	Fructose	Y	21	Adonitol	N
5	Dextrose	Y	22	Arabitol	Y
6	Galactose	Y	23	Erythritol	N
7	Raffinose	Y	24	Rhamnose	Y
8	Trehalose	Y	25	Cellobiose	Y
9	Melibiose	Y	26	Melezitose	N
10	Sucrose	Y	27	alpha-Methyl-D-Mannoside	N
11	L-Arabinose	Y	28	Xylitol	N
12	Mannose	N	29	ONPG	N
13	Inulin	Y	30	Esculin	Y
14	Sodium gluconate	N	31	D-Arabinose	Y
15	Glycerol	N	32	Citrate	Y
16	Salicin	N	33	Malonate	Y
17	Dulcitol	N	34	Sorbose	Y

Y = Yes indicates its ability to utilise carbohydrate and N = No indicates could not utilise carbohydrate.

TABLE 3 Showing Antibiogram results for the strain under study.

Sl. No	Antibiotic tested	Conc. tested	Strain D2	Sl. No	Antibiotic tested	Conc. tested	Strain D2
1	Cefpodoxime	10µg	S	13	Amikacin	30µg	S
2	Chloramphenicol	30µg	S	14	CoTrimoxazole	25µg	S
3	Vancomycin	30µg	S	15	Colistin	10µg	S
4	Streptomycin	10µg	S	16	Augmentin	30µg	S
5	Rifampicin	5µg	S	17	Netillin	30µg	S
6	Levofloxacin	5µg	I	18	Norfloxacin	10µg	S
7	Ceftriaxone	30µg	S	19	Ceftriaxone	10µg	S
8	Clindamycin	2µg	S	20	Ciprofloxacin	5µg	I
9	Augmentin	30µg	S	21	Cefotaxime	30µg	S
10	Amikacin	30µg	S	22	Gentamicin	10µg	I
11	Cefixime	5µg	S	23	Furazolidone	50µg	I
12	Tetracycline	30µg	S	24	Amoxycillin	10µg	S

S: Sensitive, I: Intermediate and R: Resistance to tested antibiotics

16

17

18 **Table 4:** General genome features of various *Kocuria rhizophila* strains.

Kocuria rhizophila	**DC2201**	**FDAARGOS 302**	**P7-4**	**TPW45**	**14ASP**	**RF**	**G2**	**D2**
Size(Mb)	2.7	2.7	2.82	2.7	2.7	2.78	2.88	2.64
GC%	71.2	71.2	70.5	70.6	70.8	70.6	70.8	70.8
Genes (total)	2,364	2,358	2,483	2,347	2,474	2,477	2,557	2,305
CDS (total)	2,264	2,300	2,359	2,254	2,416	2,414	2,502	2,253
Pseudo Genes	58	58	70	40	58	47	114	35
rRNA	3,3,3	3,3,3	2,1,2	2,1,1	3,3,2	4,5,4	4,1,1	1,1,1
tRNA	46	46	46	46	46	47	46	46
Source	Soil	Food	*Siganus doliatus*	Fresh water	*Oxalis corniculata*	Soil	Slaughter house	Human Gut
Country	-	U.S	Pacific Ocean	Malaysia	-	-	Danish	India
Publication Year	2008	2017	2011	2015	2015	2015	2016	2018
Accession No.	NC 010617.1	NZ CP022039.1	NZ AFID01000001.1	NZ JWTC01000009.1	NZ LFIY01000001.1	NZ JPWX02000010.1	NZ CZJW01000034.1	PN RK01000001.1

19

20 **Table 5:** Distribution of IS elements within the genomes, where number indicates the copy number.

IS Elements	DC2201	FDAARGOS 302	P7-4	TPW45	14ASP	RF	G2	D2
IS1595	1	1	1	1	0	1	1	0
IS481	1	1	1	1	0	1	1	1
IS5	1	1	1	1	0	1	1	1
Tn3	1	1	1	1	0	1	1	1
IS30	1	1	1	1	0	1	1	0
ISL3	1	1	1	1	0	1	1	0
ISNCY	1	1	1	1	0	1	1	0
IS110	0	0	1	1	0	1	1	0
IS3	0	0	1	1	0	1	1	0
IS1380	1	1	0	1	0	0	0	0
IS91	0	0	0	1	0	1	0	1
IS256	0	0	0	0	0	0	1	1
IS1121	0	0	0	0	1	0	0	0
IS1649	0	0	0	0	1	0	0	0
IS1652	0	0	0	0	1	0	0	0
IS21	0	0	0	1	0	0	0	0
IS5564	0	0	0	0	1	0	0	0
IS66	0	0	1	0	0	0	0	0
ISAar35	0	0	0	0	1	0	0	0
ISAcba1	0	0	0	0	1	0	0	0
ISArsp1	0	0	0	0	1	0	0	0
ISArsp6	0	0	0	0	1	0	0	0
ISAzsp1	0	0	0	0	1	0	0	0
ISBli29	0	0	0	0	1	0	0	0
ISCgl1	0	0	0	0	1	0	0	0
ISCmi2	0	0	0	0	1	0	0	0

ISGdi10	0	0	0	0	1	0	0	0
ISJs1	0	0	0	0	1	0	0	0
ISKpn27	0	0	0	0	1	0	0	0
ISKrh1	0	0	0	0	1	0	0	0
ISKrh2	0	0	0	0	1	0	0	0
ISKrh3	0	0	0	0	1	0	0	0
ISLxc2	0	0	0	0	1	0	0	0
ISLxx4	0	0	0	0	1	0	0	0
ISMav3	0	0	0	0	1	0	0	0
ISMav4	0	0	0	0	1	0	0	0
ISMsm9	0	0	0	0	1	0	0	0
ISMysp4	0	0	0	0	1	0	0	0
ISMysp5	0	0	0	0	1	0	0	0
ISNfa1	0	0	0	0	1	0	0	0
ISPfr15	0	0	0	0	1	0	0	0
ISPfr16	0	0	0	0	1	0	0	0
ISPfr17	0	0	0	0	1	0	0	0
ISPfr19	0	0	0	0	1	0	0	0
ISPfr21	0	0	0	0	1	0	0	0
ISRae1	0	0	0	0	1	0	0	0
ISRhosp6	0	0	0	0	1	0	0	0
ISSco1	0	0	0	0	1	0	0	0
ISStpr1	0	0	0	0	1	0	0	0

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES

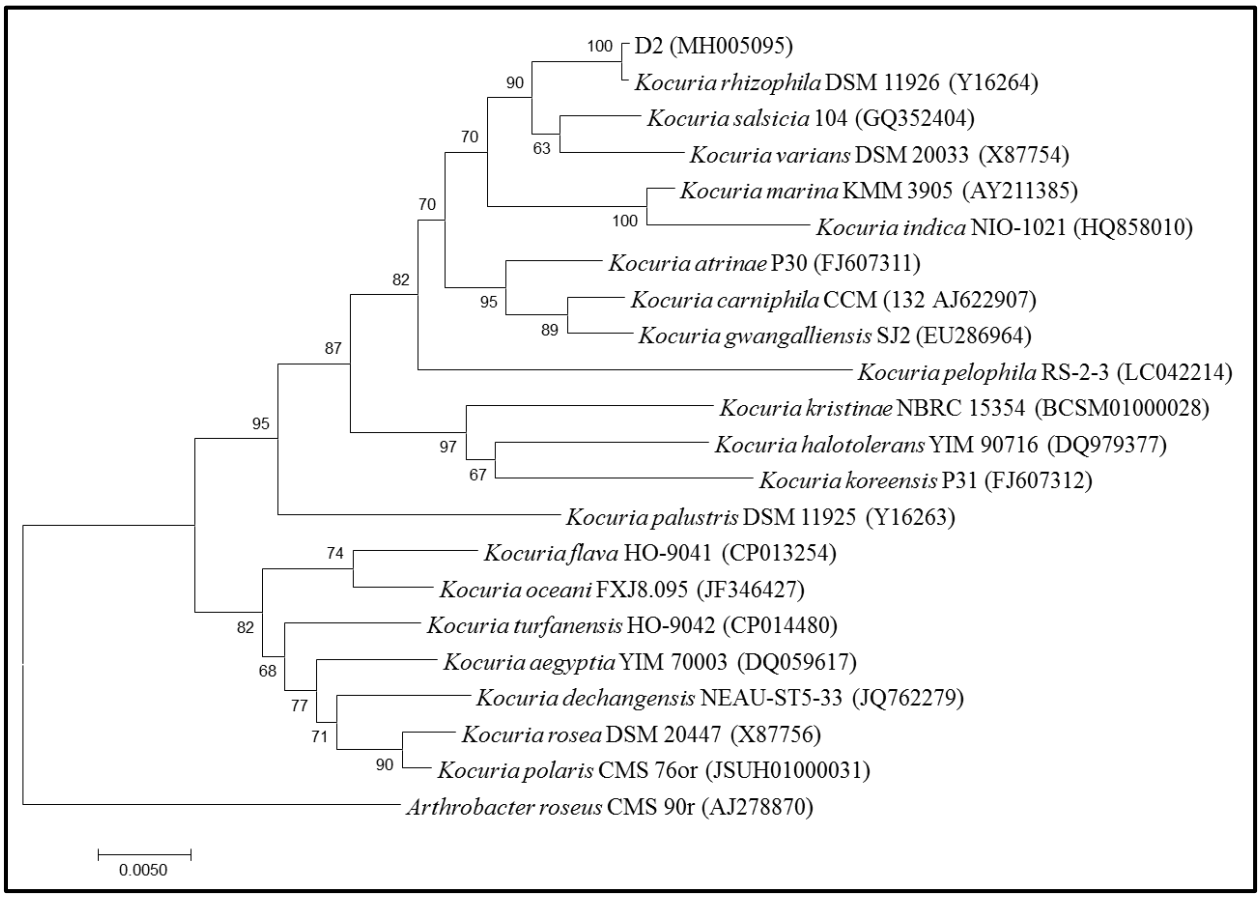

Fig 1 Phylogenetic relationship of strain D2 with closely related taxa based on 16S rRNA gene sequences (16S rRNA gene sequence of *Arthrobacter roseus* was used as out-group). The phylogenetic trees were constructed using the Neighbour-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1)

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

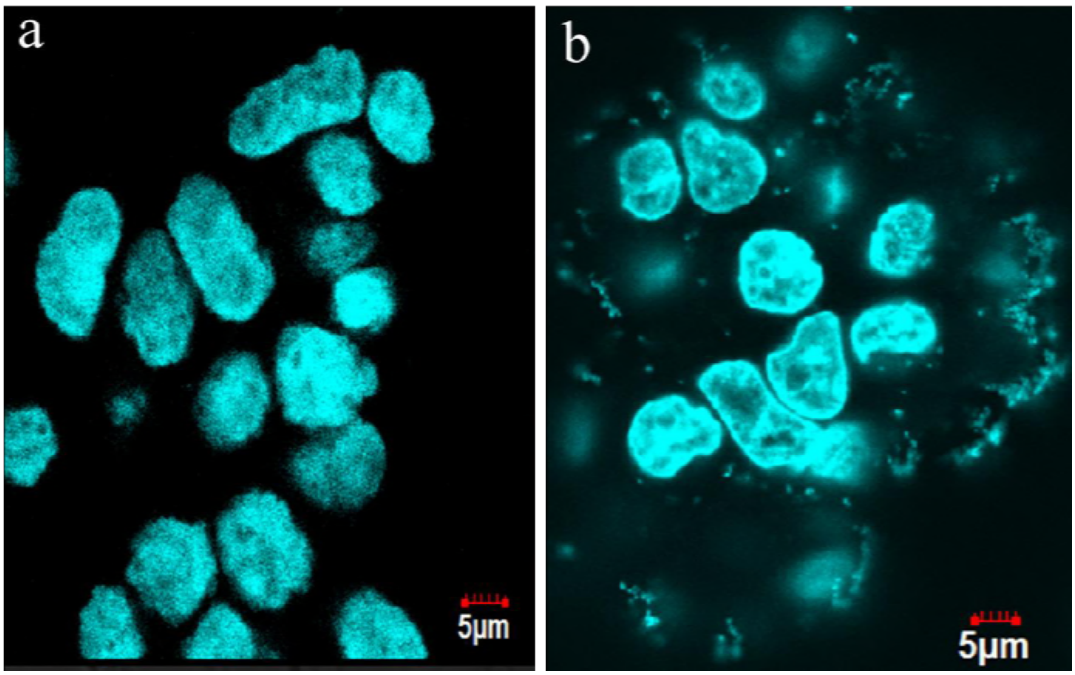

Fig 2 Results for cell adhesion assay performed on Human HT-29 Cell line. Images taken under fluorescence microscope 60X oil immersion where panel (a) is negative control without any bacterial cell added; panel (b) is for the strain D2

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

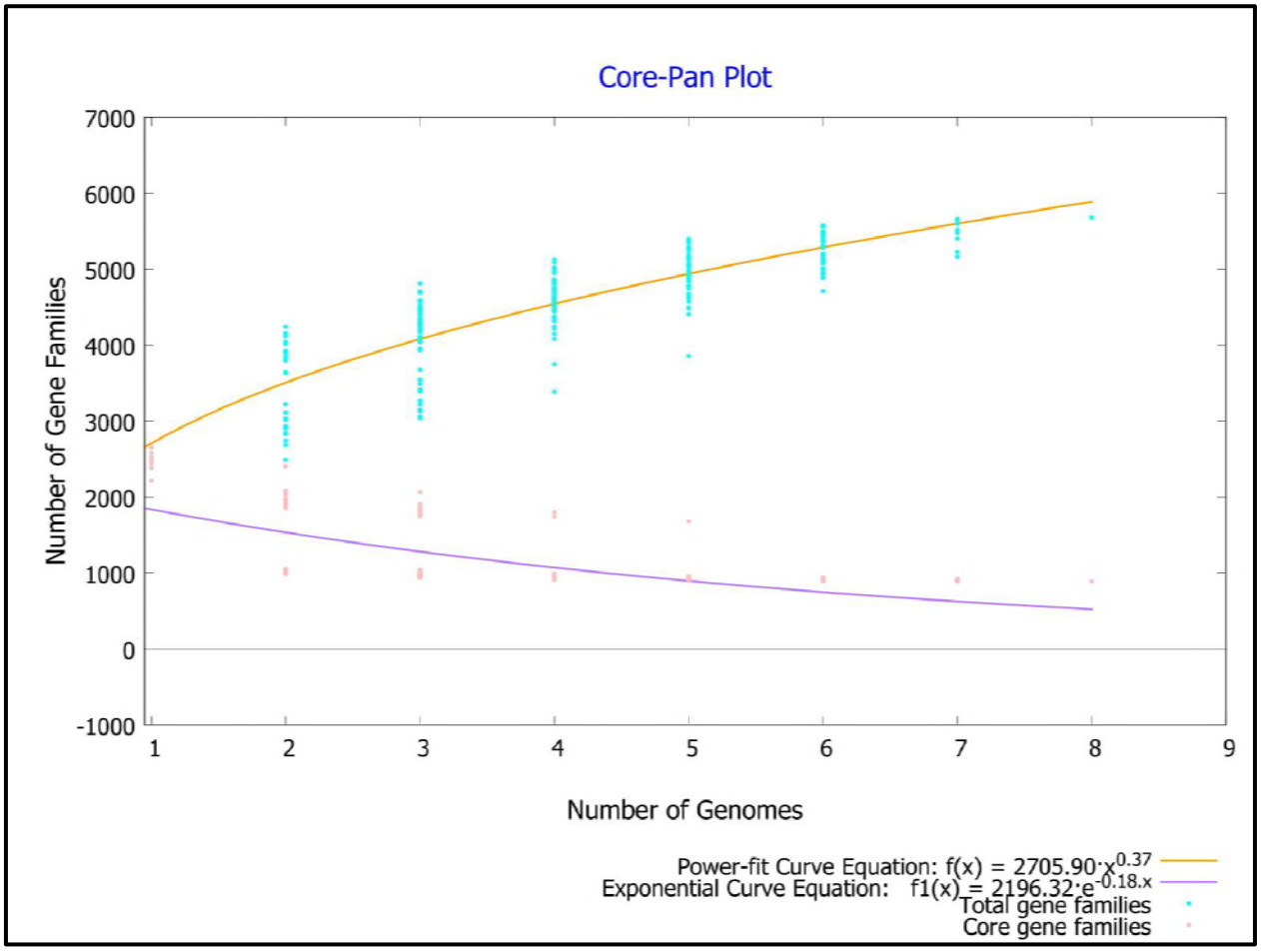

Fig 3 Core and pan genome for *K. rhizophila* strains. The number of shared genes is plotted as the function of number of strains (n) added sequentially. 888 copy genes are shared by all 8 genomes. The orange line represents the least-squares fit to the power law function f(x)=a.x^b where a= 2705.9, b= 0.373747. The red line represents the least-squares fit to the exponential decay function f1(x)=c.e^(d.x) where c= 2196.32, d= -0.179618.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

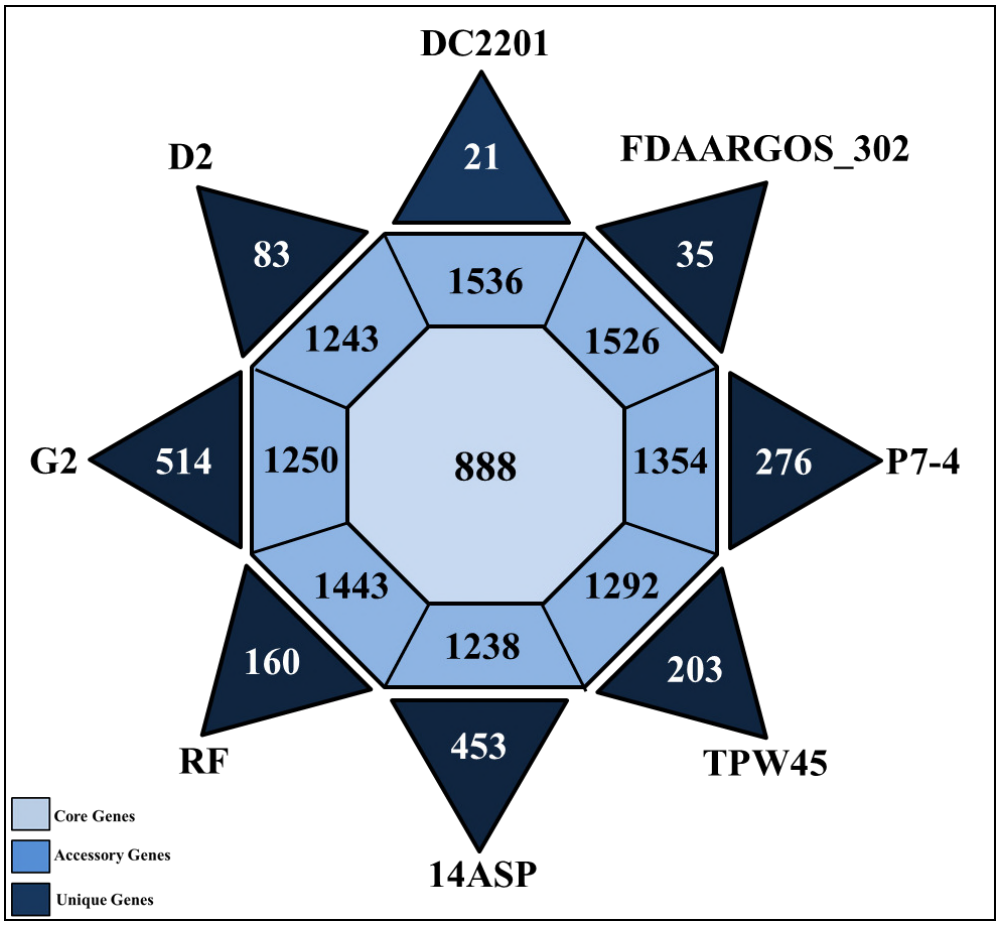

Fig 4 Number of Core, Accessory and Unique gene families of *K. rhizophila* genomes. The inner octagon represents the core genome consisting of 888 genes. The outer octagon represents the accessory genome and the number within them represents number of genes, while the triangles represent the unique genes associated with all the strains.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

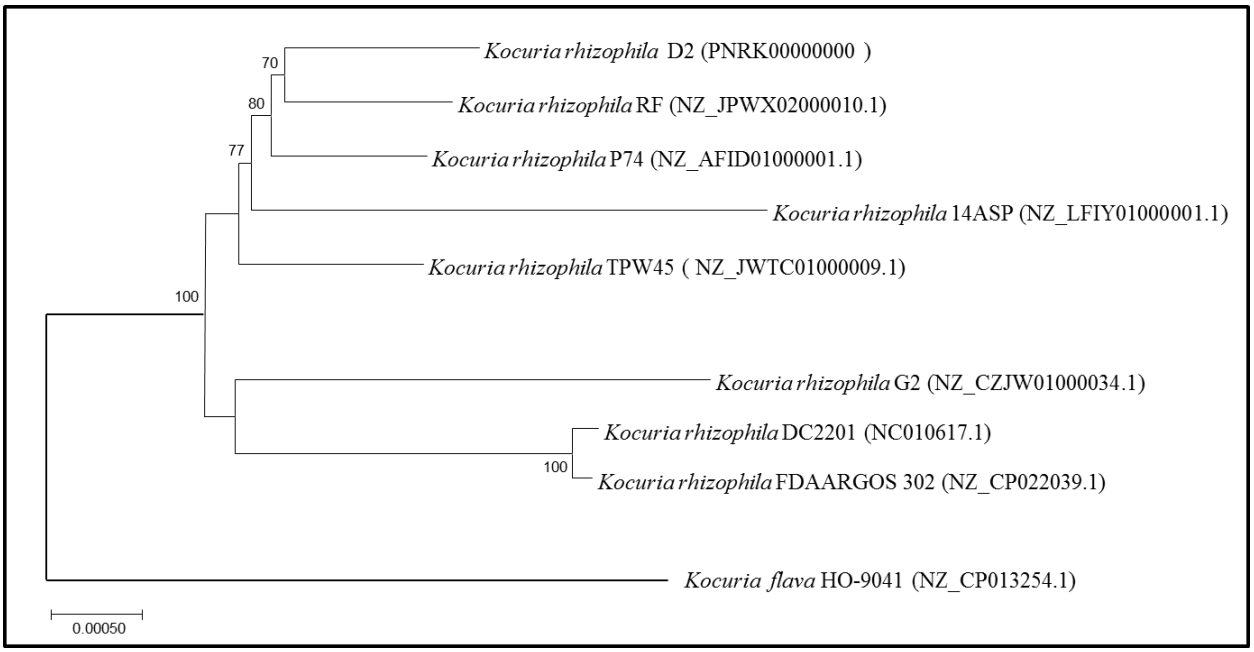

Fig 5. Core Genome Phylogeny. Phylogenetic tree of 8 *K. rhizophila* strains using the Maximum Likelihood method based on the GTR + G substitution model. The tree with the highest log likelihood (-17897.1414) is shown. Evolutionary analyses were conducted in MEGA6. A concatenated tree of 7909 genes was considered in the final dataset.

bioRxiv preprint doi: https://doi.org/10.1101/400242; this version posted August 25, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

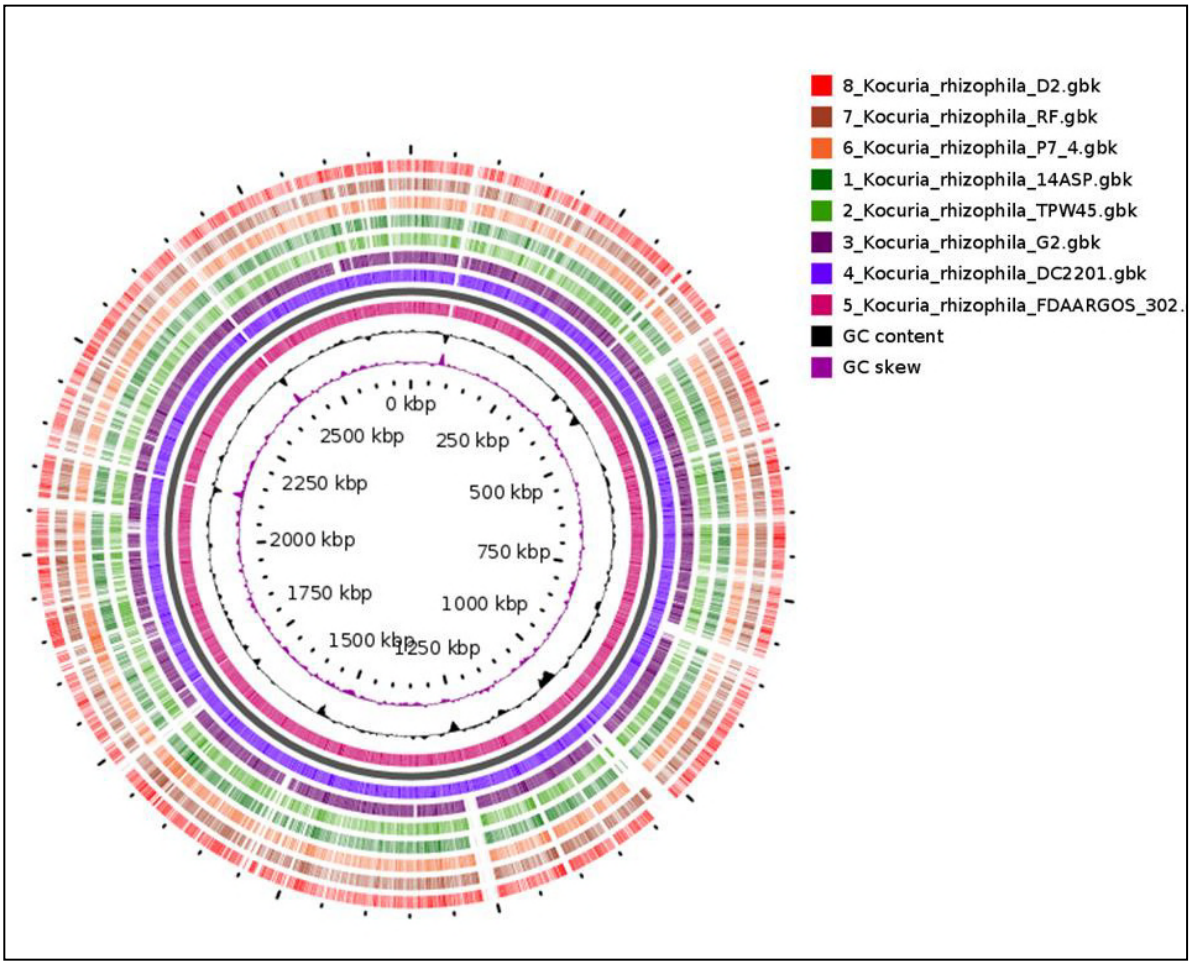

Fig 6. Blast Atlas of *Kocuria* genomes, with strain FDAARGOS_302 as a reference genome followed by DC2201, G2, TPW45, P7_4, RF and the outermost as D2. The difference between these genomes can be seen by the gaps in the rings.