



Whole Genome Sequencing and Analysis of Plant Growth Promoting Bacteria Isolated from the Rhizosphere of Plantation Crops Coconut, Cocoa and Arecanut

Alka Gupta¹*, Murali Gopal^{1*}*, George V. Thomas¹, Vinu Manikandan², John Gajewski³, George Thomas⁴, Somasekar Seshagiri⁵, Stephan C. Schuster^{3,6}, Preeti Rajesh^{2*}*, Ravi Gupta^{2*}*

1 Central Plantation Crops Research Institute, Kasaragod, Kerala, India, **2** SciGenom Labs Pvt. Ltd., Plot 43A, SDF 3rd Floor CSEZ, Kakkanad, Cochin, Kerala, India, **3** Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, 310 Wartik Lab, University Park, Pennsylvania, United States of America, **4** SciGenom Research Foundation, Cochin, Kerala, India, **5** Department of Molecular Biology, Genentech Inc., South San Francisco, California, United States of America, **6** Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technical University, Singapore, Singapore

Abstract

Coconut, cocoa and arecanut are commercial plantation crops that play a vital role in the Indian economy while sustaining the livelihood of more than 10 million Indians. According to 2012 Food and Agricultural organization's report, India is the third largest producer of coconut and it dominates the production of arecanut worldwide. In this study, three Plant Growth Promoting Rhizobacteria (PGPR) from coconut (CPCRI-1), cocoa (CPCRI-2) and arecanut (CPCRI-3) characterized for the PGP activities have been sequenced. The draft genome sizes were 4.7 Mb (56% GC), 5.9 Mb (63.6% GC) and 5.1 Mb (54.8% GB) for CPCRI-1, CPCRI-2, CPCRI-3, respectively. These genomes encoded 4056 (CPCRI-1), 4637 (CPCRI-2) and 4286 (CPCRI-3) protein-coding genes. Phylogenetic analysis revealed that both CPCRI-1 and CPCRI-3 belonged to *Enterobacteriaceae* family, while, CPCRI-2 was a *Pseudomonadaceae* family member. Functional annotation of the genes predicted that all three bacteria encoded genes needed for mineral phosphate solubilization, siderophores, acetoin, butanediol, 1-aminocyclopropane-1-carboxylate (ACC) deaminase, chitinase, phenazine, 4-hydroxybenzoate, trehalose and quorum sensing molecules supportive of the plant growth promoting traits observed in the course of their isolation and characterization. Additionally, in all the three CPCRI PGPRs, we identified genes involved in synthesis of hydrogen sulfide (H₂S), which recently has been proposed to aid plant growth. The PGPRs also carried genes for central carbohydrate metabolism indicating that the bacteria can efficiently utilize the root exudates and other organic materials as energy source. Genes for production of peroxidases, catalases and superoxide dismutases that confer resistance to oxidative stresses in plants were identified. Besides these, genes for heat shock tolerance, cold shock tolerance and glycine-betaine production that enable bacteria to survive abiotic stress were also identified.

Citation: Gupta A, Gopal M, Thomas GV, Manikandan V, Gajewski J, et al. (2014) Whole Genome Sequencing and Analysis of Plant Growth Promoting Bacteria Isolated from the Rhizosphere of Plantation Crops Coconut, Cocoa and Arecanut. PLoS ONE 9(8): e104259. doi:10.1371/journal.pone.0104259

Editor: Matteo Pellegrini, UCLA-DOE Institute for Genomics and Proteomics, United States of America

Received: November 28, 2013; **Accepted:** July 9, 2014; **Published:** August 27, 2014

Copyright: © 2014 Gupta et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was internally funded by SciGenom Labs Pvt Ltd. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Some of the authors as noted are employees of SciGenom Pvt Ltd. S.S. is an employee of Genentech and holds shares in Roche. The study was internally funded by SciGenom Labs Pvt Ltd. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

* Email: mgpcpri@yahoo.co.in (MG); preeti@scigenom.com (PR); ravig@scigenom.com (RG)

† These authors contributed equally to this work.

Introduction

Plant rhizosphere harbors numerous bacteria capable of stimulating and aiding plant growth and are termed plant growth promoting rhizobacteria (PGPR) [1]. They exert their beneficial effects through direct or indirect mechanisms. The direct mechanisms include biofertilization, stimulation of root growth, rhizo-remediation and plant stress control [2]. Indirect mechanisms primarily involve biological control comprised of antibiosis, induction of systemic resistance and competition for nutrition and niches [2]. Owing to their diverse plant growth promoting capabilities, PGPRs have become the new inoculants for biofertilizer technology [3]. To improve the biofertilizer technology, understanding the molecular mechanisms of plant growth

promotion and biocontrol by rhizobacteria is important [4]. Identification of genes that contribute to the beneficial activity of rhizobacteria, besides adding to our understanding of the molecular mechanisms, will aid in developing better biofertilizers.

Next generation sequencing technologies (NGS) have enabled whole genome sequencing of bacteria and other organisms [5]. Systematic analysis of whole genome data has aided the understanding of the molecular genetics of many bacterial species [6]. Recently, NGS has been employed to study genomes of several PGPRs, mainly isolated from crop species such as wheat [7], *Miscanthus* [8], pepper [9]. PGPRs from soil have also been sequenced directly [10]. However, thus far, genome sequences of

PGPRs isolated from plantation crops, particularly from coconut, cocoa and arecanut, have not been reported.

Coconut (*Cocos nucifera* L.), cocoa (*Theobroma cacao* L.) and arecanut (*Areca catechu* L.) are three important plantation crops grown in 2.2 million hectares in India. These plantation crops harbor plant-beneficial microorganisms in their rhizospheres [11–15] and some of these have been utilized for growth promotion [16–20] as they offer an opportunity for ecologically safe nutrient management [21]. In this study, we have performed deep sequencing analysis of three PGPRs, CPCRI-1, CPCRI-2 and CPCRI-3, isolated from coconut [16], cocoa [20] and arecanut [22] rhizosphere, respectively.

Results

PGPR strains

Soil samples collected from rhizospheres of coconut, cocoa and arecanut grown in different agro-ecological zones of India were used to isolate 1512 morphologically distinct heterotrophic bacteria [13,15,18,22]. The details of places from which the soil samples were collected, soil types and their pH along with isolation media are given in Table S1. The isolates were screened *in vitro* for several important plant growth promoting functions (Table 1). The isolates that gave best results in the *in vitro* testing were then studied for plant growth promotion using rice and cowpea seeds in environmental growth chamber and green house conditions [18,19]. They were also then tested on coconut [16], cocoa [20] and arecanut [22] seedlings grown in polybags.

Three PGPRs designated CPCRI-1 (RNF-267 from coconut) [16], CPCRI-2 (KGSF-20 from cocoa) [15,20] and CPCRI-3 (KtRA5-88 from arecanut) [22] were selected for further studies. All the three isolates had rod shape morphology and were negative for Gram's staining. CPCRI-1 showed good phosphate solubilizing capacity and promoted growth of coconut seedlings [16]. CPCRI-2 was capable of promoting growth of cocoa seedlings [20]. CPCRI-3, isolated from arecanut rhizosphere, was able to tolerate low pH and possessed plant growth promoting attributes [22]. The plant growth promotion traits of the three isolates are summarized in Table 1. The morphological, biochemical and physiological attributes of the three PGPRs are summarized in

Table S2. Given the beneficial attributes of the three PGPRs, we chose to characterize them further at the genomic level.

Whole genome shotgun sequencing and assembly

We performed shotgun multiplexed sequencing of the genomes of CPCRI-1, CPCRI-2 and CPCRI-3 using the 454-sequencing platform. We obtained >300,000 quality-filtered reads each for CPCRI-1 and CPCRI-3 with an average read length of 465 bp and 421 bp, respectively. For CPCRI-2, we obtained >150,000 quality filtered reads with an average read length of 408 bp (Table 2). We assembled the sequencing reads for each of the three genomes using *GS de novo* assembler version 2.6 [23]. Of the total reads obtained ~90% were assembled into contigs corresponding to each of the genomes. For CPCRI-1, 350,636 reads were assembled into 39 contigs (N50 of 242,562 bp; longest contig length of 730,806 bp; mean contig length of 114,755 bp) for a total of 4,475,442 nucleotides at ~30× coverage. The 144,293 reads obtained for CPCRI-2, were assembled into 101 contigs (N50 of 89,849 bp; longest contig length of 282,342 bp; mean contig length of 52,329 bp) for a total of 5,285,206 nucleotides at ~12× coverage. In the case of CPCRI-3, the 313,271 reads obtained were assembled into 47 contigs (N50 of 161,752 bp; longest contig length of 529,776 bp; mean contig length of 99,348 bp) for a total of 4,669,355 nucleotides at ~30× coverage (Table 2, Fig. S1).

The estimated genome size based on the sequence data was 4.7 Mb for CPCRI-1, 5.9 Mb for CPCRI-2 and 5.1 Mb for CPCRI-3. Phylogenetic analysis derived from comparison of 31 conserved housekeeping protein-coding genes [24] indicated that while CPCRI-1 and CPCRI-3 were members of the *Enterobacteriaceae* family, CPCRI-2 was a *Pseudomonadaceae* family member. Their estimated genome sizes are consistent with the sizes observed for other family members (Table S3 & S4). The GC content of the bacterial isolates was 56.0%, 63.6% and 54.8% for CPCRI-1, CPCRI-2 and CPCRI-3, respectively (Table 2).

Gene prediction and annotation

Glimmer-MG [25] predicted 4056, 4637 and 4286 protein-coding genes in CPCRI-1, CPCRI-2 and CPCRI-3, respectively (Table 3 & Table S5). Consistent with this the average predicted

Table 1. Biological and plant growth promotional properties of PGPR isolates.

S. No.	Attributes	CPCRI-1 (RNF 267) [16]	CPCRI-2 (KGSF20) [20]	CPCRI-3 (KtRA5-88) [22]
1.	pH tolerance levels	4.0 to 9.0	4.0 to 8.0	2.5 to 7.0
2.	Optimum pH for growth	5.0–6.0	7.0	4.5
3.	NaCl tolerance	upto 8%	upto 4%	upto 2%
4.	Temperature tolerance	15 to 40°C	15 to 40°C	30 to 40°C
5.	ACC deaminase activity	562 $\mu\text{mol } \alpha\text{-ketobutyrate h}^{-1} \text{ mg protein}^{-1}$	199 $\mu\text{mol } \alpha\text{-ketobutyrate h}^{-1} \text{ mg protein}^{-1}$	474 $\mu\text{mol } \alpha\text{-ketobutyrate h}^{-1} \text{ mg protein}^{-1}$
6.	Phosphate solubilization	217 $\mu\text{g ml}^{-1}$	99.8 $\mu\text{g ml}^{-1}$	82 $\mu\text{g ml}^{-1}$
7.	IAA production	2.4 $\mu\text{g ml}^{-1}$	1.5 $\mu\text{g ml}^{-1}$	1.2 $\mu\text{g ml}^{-1}$
8.	Growth on N-free agar medium	–	Growth observed	–
9.	Chitinase activity	–	–	–
10.	β -1,3-glucanase activity	20 $\mu\text{g glucose min}^{-1} \text{ mg protein}^{-1}$	7.8 $\mu\text{g glucose min}^{-1} \text{ mg protein}^{-1}$	2.4 $\mu\text{g glucose min}^{-1} \text{ mg protein}^{-1}$
11.	Salicylic acid production	–	6.1 $\mu\text{g ml}^{-1}$	–
12.	Siderophore production	–	6 mm	11 mm

– indicates no growth/production/activity.

doi:10.1371/journal.pone.0104259.t001

Table 2. Genome assembly statistics.

	CPCRI-1	CPCRI-2	CPCRI-3
# of reads	361,881	158,071	326,453
Average read length (bp)	465	408	421
# of bases (bp)	168,202,148	64,520,437	137,575,981
# of reads assembled	350,636 (96.9%)	144,293 (91.3%)	313,271 (96%)
Number of contigs (> = 500 bp)	39	101	47
N50 (bp)	242,562	89,849	161,752
Average contig length (bp)	114,755	52,329	99,348
Total length of contigs (bp)	4,475,442	5,285,206	4,669,355
GC Content (%)	56.0	63.6	54.8
Average contig size (bp)	114,755	52,329	99,348
Longest contig size (bp)	730,806	282,342	529,776
Q40 plus bases (%)	99.97	99.77	99.96

doi:10.1371/journal.pone.0104259.t002

protein coding genes size in CPCRI-1, CPCRI-2 and CPCRI-3 was found to be 972 bp, 981 bp and 951 bp, respectively. In bacteria, a robust correlation exists between the genome size and the numbers of genes it encodes [26]. A comparison of 26 published complete genomes in the *Enterobacteriaceae* family revealed an average genome size of 4.8 Mb and the coded for an average of 4655 proteins (Table S3). Our estimate of 4056 genes in CPCRI-1 and 4286 genes in CPCRI-3 is consistent with this observation. The *Pseudomonas* genus had an average genome size of 6.0 Mb and encoded an average of 5366 protein coding genes (Table S4). Though CPCRI-2 had a genome size of 5.9 Mb, and coded for about 4637 proteins, this number is similar to those observed in *Pseudomonas putida* BIRD-1, a PGPR [10].

The average GC content of the protein coding genes in CPCRI-1 (56.55%), CPCRI-2 (64.12%) and CPCRI-3 (55.4%) and their relation with the genomic GC content were estimated and are presented in Table 3, Table S3, S4. The GC distribution for all three positions in the codon for protein-coding genes (Fig. S2) showed that the average GC content was the highest for the third base position and lowest for second position within the codon. Though the overall trend was similar between the bacteria, the GC content for CPCRI-2 at the third base was ~85% compared to only 68% for CPCRI-1 and CPCRI-3. The codon usage

analysis (Fig. S3) showed that CTG that codes for Leu (L) is the most used codon in all three bacteria.

In addition to the protein coding genes, we predicted 74, 61, and 74 tRNA genes in CPCRI-1, CPCRI-2, and CPCRI-3, respectively (Table S6A, B, C). The predicted genes represented 20–21 different tRNAs corresponding to the universal codons. Our analysis additionally identified 5–6 different rRNA genes (Table 3 & Table S7).

To further understand the bacterial strains, the predicted protein-coding genes identified using Glimmer-MG were compared against the non-redundant (nr) NCBI protein database using BLASTX [27]. We found that a majority of the predicted protein-coding genes (98%) had a homologous protein sequence in the NCBI non-redundant (nr) protein database (Fig. 1). Among the genes with homologs, >90% of genes had a high confidence match ($E\text{-value} \leq 1.0 \times 10^{-50}$) and >93% had identity of at least 80% with a putative homolog (Fig. 1A, B). Interestingly, 31, 59 and 91 genes from CPCRI-1, 2 and 3, respectively, showed no significant identity to sequences in the NCBI database. Of the protein with no significant identity, we annotated protein domains for 15, 21 and 29 genes from CPCRI-1, 2 and 3, respectively, using InterPro [28] and CDD [29]. We further annotated all the

Table 3. Gene prediction and annotation summary.

	CPCRI-1	CPCRI-2	CPCRI-3
# of predicted protein-coding gene	4,056	4,637	4,286
Average GC content of Protein-coding genes (%)	56.55	64.12	55.4
Average gene length (bp)	972	981	951
# of tRNA genes	74	61	74
Average length of tRNA gene (bp)	76	75	76
# of rRNA genes	5	6	6
# of protein-coding genes with at least one significant BLASTX result	4,053	4,596	4,216
# of BLASTX protein hit present in UniProt	3,747	4,551	3,674
# of protein-coding genes assigned GO term	2,906	3,379	2,515
# of protein-coding genes assigned an InterPro id	3,466	4,097	3,063

doi:10.1371/journal.pone.0104259.t003

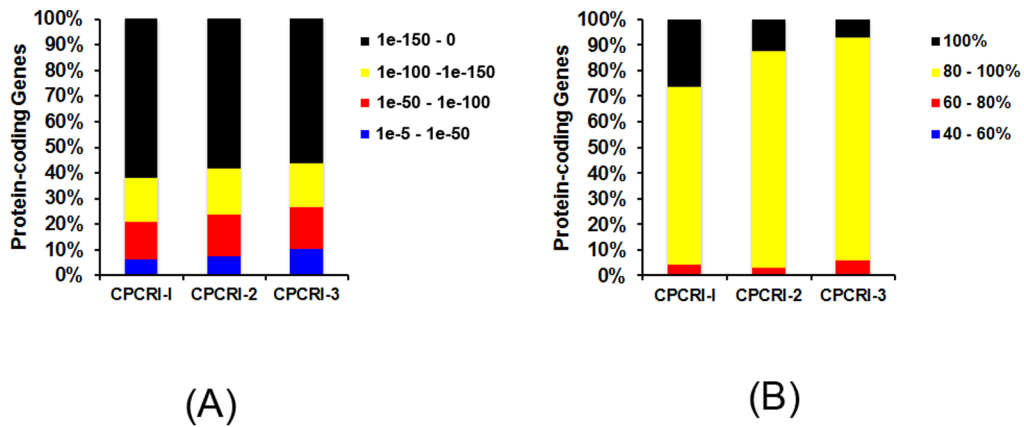


Figure 1. Protein coding genes. Stacked bar graph representing the percentage of predicted protein coding genes with significant matches ($E\text{-value} \leq 1e10^{-5}$) (A) in the NCBI nr-protein database identified using BLASTX and (B) the proportion of proteins binned by percent identity measured by BLASTX alignment.
doi:10.1371/journal.pone.0104259.g001

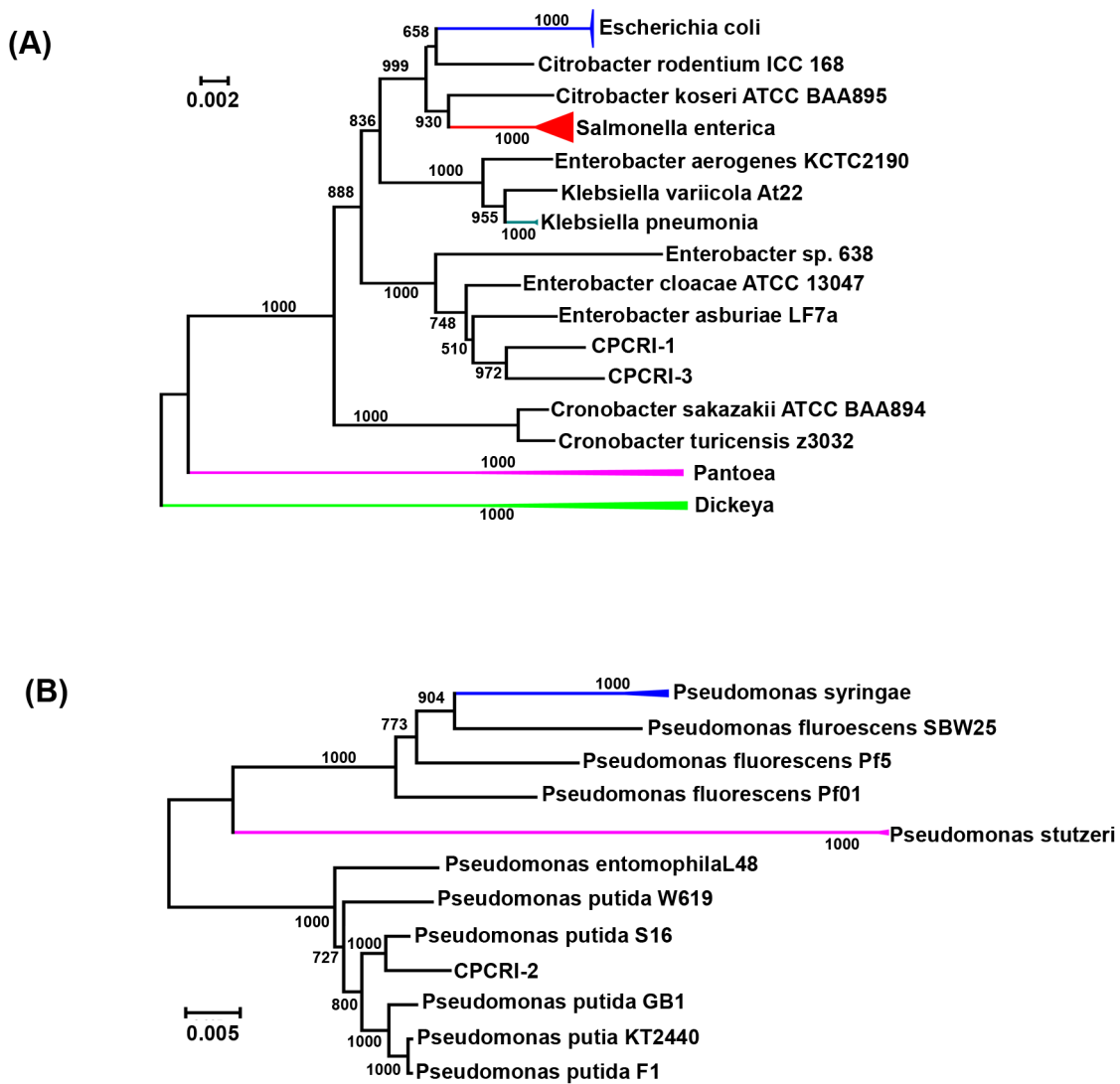


Figure 2. Phylogenetic tree. Using 31 conserved housekeeping protein-coding genes from (A) CPCRI-1 and CPCRI-3, (B) CPCRI-2, a phylogenetic tree was generated using AMPHORA2 [24,94] and ClustalW [95]. The colored branch/node represents node where multiple strains of the same species are collapsed into a single species for representation.
doi:10.1371/journal.pone.0104259.g002

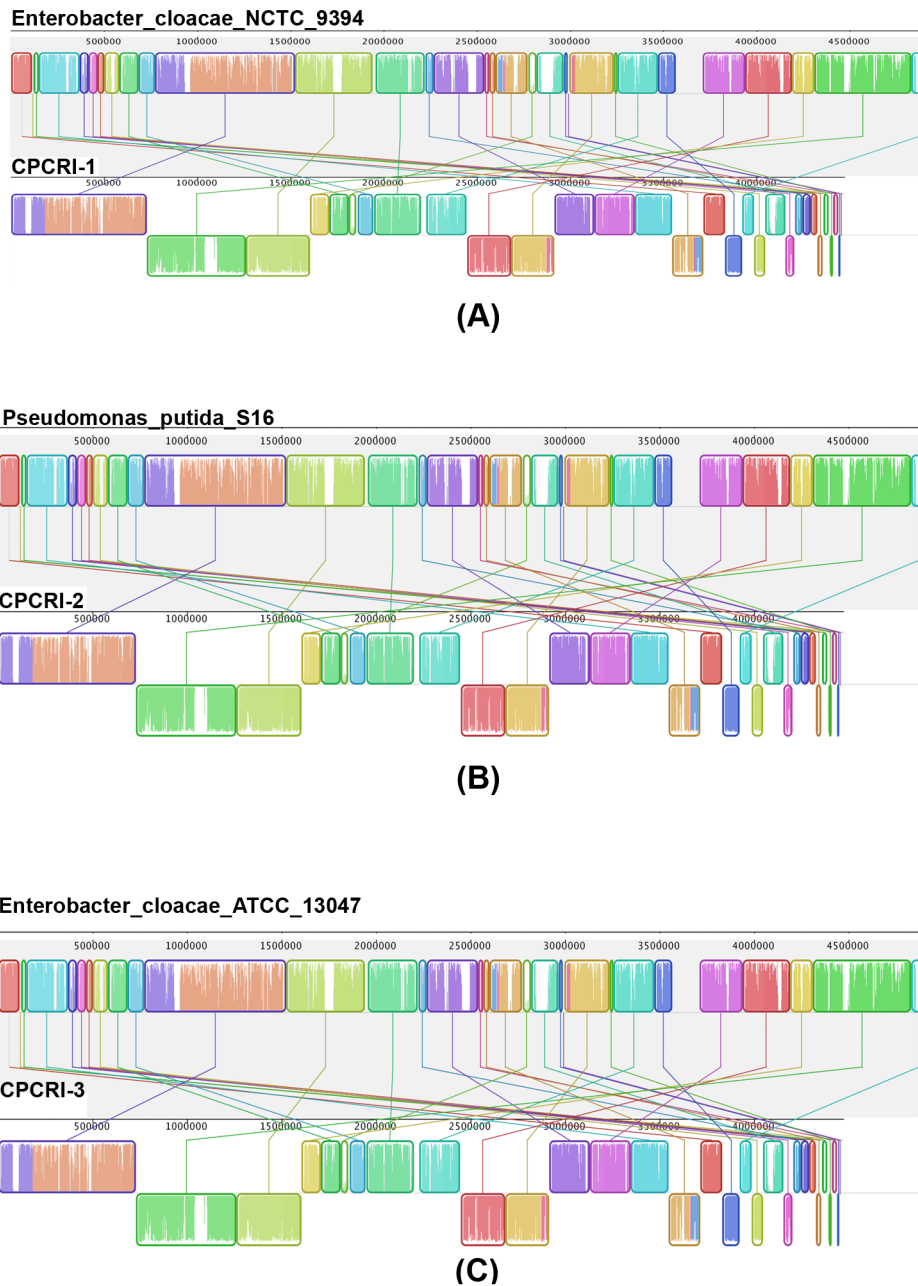


Figure 3. Genome comparison. Pairwise alignment of CPCRI-1, CPCRI-2, CPCRI-3 genome with *Enterobacter cloacae* NCTC 9394, *Pseudomonas putida* S16 and *Enterobacter cloacae* ATCC 13047, respectively using the progressive Mauve aligner [34]. The colored blocks represent the homologous region between the genomes that are internally free from genomic rearrangement. doi:10.1371/journal.pone.0104259.g003

protein-coding genes with known protein domains using UniProt database (Table S8A, B, and C) [30].

Phylogeny analysis

The phylogenetic analysis performed using a set of 31 conserved housekeeping protein-coding genes [24] revealed that the CPCRI-1 and CPCRI-3 genomes were closely related to the *Enterobacter cloacae* group (Fig. 2A). The CPCRI-2 genome was found to be closely related to the *Pseudomonas putida* group. CPCRI-1 grouped with *Enterobacter asburiae* strain LF7a and *Enterobacter cloacae* ATCC 13047, both of which are members of the *Enterobacter cloacae* complex. CPCRI-3 clustered closely with

Enterobacter asburiae strain LF7a and *Enterobacter* sp. 638 [31], an endophyte of poplar trees. The closest relative to CPCRI-2 was *Pseudomonas putida* strain S16, a gram-negative soil bacterium with an ability to degrade aromatic and heterocyclic compounds, such as nicotine, benzoate, and phenylalanine [32]. Taxonomy based study using MEGAN4 [33] showed that CPCRI-1 and CPCRI-3 belong to *Enterobacteriaceae* family and CPCRI-2 belong to *Pseudomonas* genus (Fig. S4, Result S1). Consistent with this, Biolog analysis indicated CPCRI-2 to be *Pseudomonas putida* [20]. Although Biolog analysis at low confidence level, indicated CPCRI-3 to be *Pantoea agglomerans* [22], a member of

Table 4. Pairwise comparison of CPCRI-1, CPCRI-2, and CPCRI-3 genomes against bacteria genomes using progressive Mauve aligner [34].

Bacteria	CPCRI-1	CPCRI-2	CPCRI-3
<i>Enterobacter asburiae</i> LF7a	84.94/83.15	65.25/14.5	81.26/78.61
<i>Enterobacter cloacae</i> ATCC 13047	86.20/87.75	65.88/14.67	81.58/78.56*
<i>Enterobacter cloacae</i> EcWSU1	85.86/86.70	65.38/14.88	81.4/77.95
<i>Enterobacter cloacae</i> NCTC 9394	92.69/89.02*	65.98/14.47	81.4/77.38
<i>Enterobacter cloacae</i> SCF1	79.12/72.00	66.85/12.69	78.3/66.83
<i>Enterobacter</i> sp. 638	82.0/78.73	64.79/13.36	80.57/75.66
<i>Pseudomonas putida</i> BIRD-1	65.60/18.09	87.84/81.05	65.19/17.06
<i>Pseudomonas putida</i> F1	65.50/18.94	88.05/83.70	65.2/17.04
<i>Pseudomonas putida</i> GB-1	65.47/18.7	87.7/82.97	65.16/17.57
<i>Pseudomonas putida</i> KT2440	65.57/18.14	87.69/83.0	65.21/17.0
<i>Pseudomonas putida</i> S16	65.91/18.01	89.88/85.0*	65.31/16.91
<i>Pseudomonas putida</i> W619	65.50/18.46	84.67/78.64	65.14/16.84
<i>Pseudomonas aeruginosa</i> LESB58	65.87/15.85	75.35/59.15	65.56/15.53
<i>Pseudomonas fluorescens</i> F113	65.43/18.89	76.3/66.19	64.93/17.05
<i>Pseudomonas syringae</i> pv tomato str DC3000	65.26/17.91	74.91/55.95	64.6/16.93

Percentage sequence similarity/genome coverage in the conserved block is shown.

*Bacteria showing highest similarity/genome coverage with sequenced bacteria CPCRI-1, CPCRI-2 and CPCRI-3.

doi:10.1371/journal.pone.0104259.t004

the *Enterobacteriaceae*, the genome sequence of CPCRI-1 did not reveal similarity to any sequenced bacterium at the species level.

Pairwise genome comparison with existing bacterial genomes

We performed a pairwise genome comparison of our assembled bacterial genomes against 40 different bacteria using progressive Mauve aligner [34]. The bacterial groups identified for analysis included *Enterobacter*, *Escherichia coli*, *Pseudomonas putida*, *Citrobacter*, *Dickeya*, *Klebsiella*, *Pantoea*, *Salmonella*, *Shigella*, *Azotobacter*, *Bradyrhizobium*, *Mesorhizobium* and *Rhizobium*. The genome level comparison showed that CPCRI-1 had the highest similarity to *Enterobacter cloacae* NCTC 9394i (similarity score of 92.69%, coverage of 89.02%). In addition, CPCRI-2 was closest to *Pseudomonas putida* S16 (similarity score of 89.88%, coverage of 85.0%), and CPCRI-3 to be most similar to *Enterobacter cloacae* ATCC 13047 (similarity score of 81.58%, coverage of 78.56%; Table 4, Table S9 and Fig. 3). These results are consistent with the phylogenetic analysis findings that showed CPCRI-1 and CPCRI-3 belong to *Enterobacter cloacae* group and CPCRI-2 to the *Pseudomonas putida* group.

Functional analysis of the bacterial genome

We performed functional analysis of the annotated genomes using gene ontology (GO), SEED classification and KEGG pathways. The GO based classification of the genes revealed 2,200 to 3,000 (2,562 for CPCRI-1, 3,020 for CPCRI-2, and 2,226 for CPCRI-3) genes associated with at least one molecular function, 1,500–1,900 (1,836 for CPCRI-1, 1,918 for CPCRI-2, and 1,555 for CPCRI-3) genes associated with at least one biological process, and 1,200–1,500 (1,487 for CPCRI-1, 1,530 for CPCRI-2, and 1,226 for CPCRI-3) genes associated with at least one cellular component (Table S10). The Carbohydrate metabolism, chemotaxis, cell adhesion, cilium or flagellum-dependent related motility, response to stress, iron ion binding, oxidoreductase activity are among the top 20 GO biological processes and molecular

functions found in the bacteria (Fig. 4A, B). These terms are related to the functional class of genes that aid the plant growth.

The SEED based classification [35] analysis of the proteins, performed using MEGAN4 [33] assigned functional roles to the annotated genes that were then grouped into one or more subsystems. MEGAN4 classified 1998, 2202 and 2030 annotated genes from CPCRI-1, CPCRI-2 and CPCRI-3, respectively into 25 functional categories (Fig. 4C). A large number of genes fall into carbohydrate metabolism, stress response, motility and chemotaxis and metabolism of aromatic compounds that helps in plant growth. Annotation against KEGG pathway classified the protein-coding genes into six different pathway categories: metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems and human diseases (Fig. 4D). The metabolism and environment information processing category pathways were highly represented in all three bacteria. Overall we found that many genes fall into the functional classes that support plant growth.

Plant growth promoting properties

In the genomic sequence of three PGPRs sequenced we identified genes that can be attributed to their ability to improve nutrient availability, suppress pathogenic fungi, resist oxidative stress, quorum sensing and ability to break down aromatic and toxic compounds and other abiotic stress (Table 5). The genomes of CPCRI-1, 2 and 3 possessed genes encoding glucose dehydrogenase activity while, CPCRI-1 and 2 carried the cofactor pyrrolo-quinolone quinine (*pqq*) gene cluster which is involved in solubilization of mineral phosphates fixed in soil particles [36]. Indole acetic acid (IAA) is an important hormone that helps in plant growth [37]. Although IAA production was observed in culture from the three PGPRs was low (Table 1), CPCRI-1 and CPCRI-3 genomes contained *ipdC* that codes for indolepyruvate decarboxylase, an enzyme that produces indole acetic acid from tryptophan [38]. In these genomes we also found some of the trp cluster (*trpA*, B, D, C, R) genes involved in

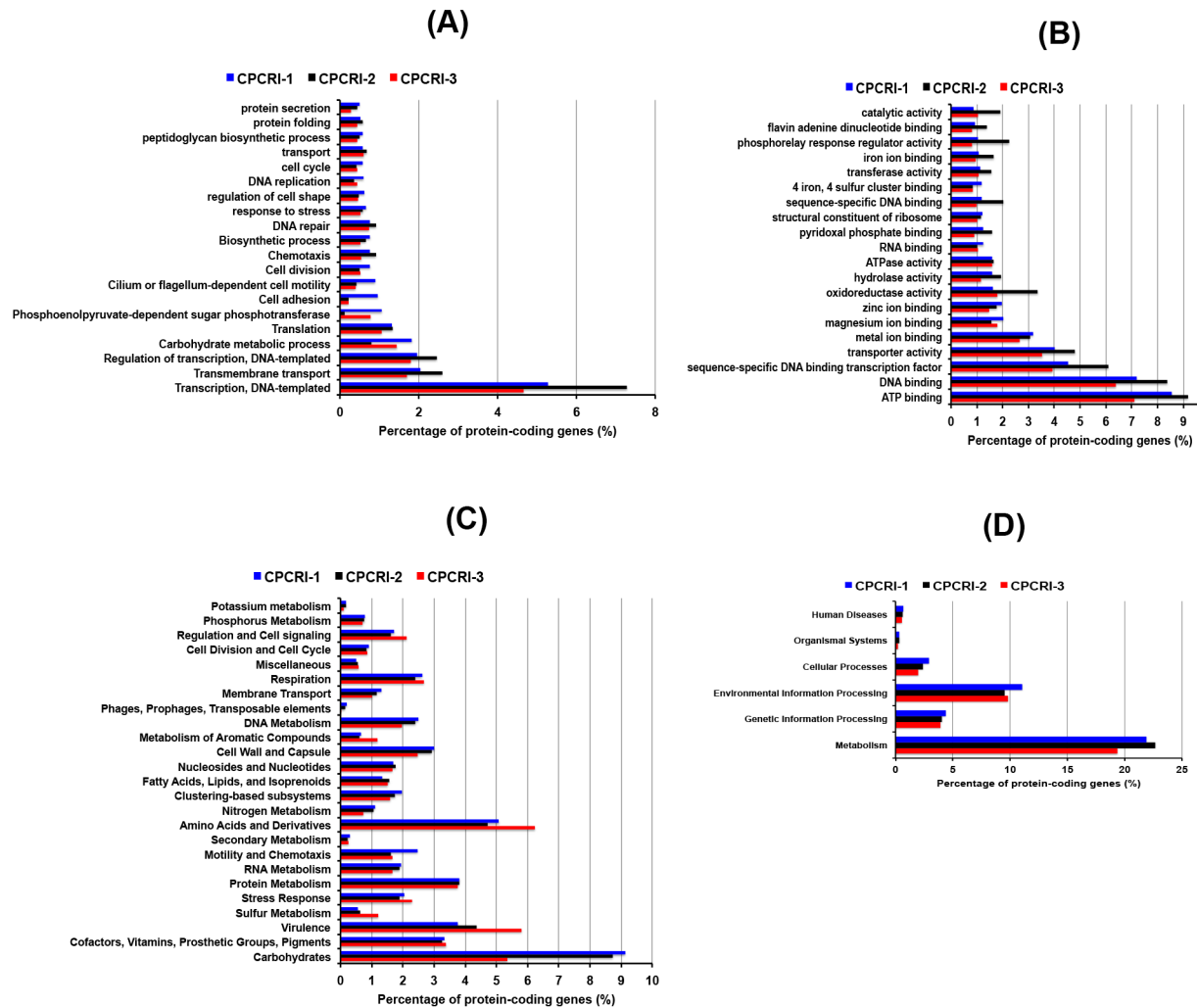


Figure 4. Functional annotation. The percentage of predicted coding genes of CPCRI PGPR genomes into (A) GO biological classes, (B) GO molecular classes, (C) SEED classification, (D) KEGG pathway classification. doi:10.1371/journal.pone.0104259.g004

tryptophan biosynthesis. These genes may play a role in synthesis of tryptophan used in multiple biological processes, including IAA biosynthesis. The 1-aminocyclopropane-1-carboxylate (ACC) deaminase has been shown in symbiotic bacteria to function in lowering the plant ethylene known to inhibit the nodulation process [39]. We identified in CPCRI-2 *acdS* gene homologue that codes for ACC deaminase. In CPCRI-1 and 3, we found, *rimM* [37] and *dcyD* [40], both of which also code for ACC deaminase. In addition, we found genes involved in hydrogen sulfide (H_2S) biosynthesis in all the three PGPR genomes (CPCRI-1, Gene#2608; CPCRI-2, Gene#3152; CPCRI-3, Gene#3465-67, 3470-72). Recently, H_2S has been reported to increase plant growth and seed germination [41] and the H_2S production by the PGPRs may play an analogous role in plant roots they colonize.

Among recently described volatile molecules known to directly influence the plant growth promotion are acetoin and 2,3-butanediol [42]. The CPCRI genomes encode *budC*, *budA* [43] and *als* [44], all of which are involved in the production of acetoin and 2,3-butanediol.

Many PGPRs are known to have biocontrol activities. Production and secretion of siderophores widely used by bacteria for iron acquisition is one of the modes of biocontrol activity.

While CPCRI-2 encoded 41 genes involved in the production and utilization of pyoverdine, siderophore, CPCRI-1 and 3, have additional genes such as *fpv* and *mbt* [45] that are linked to pyoverdine production. Besides the genes for pyoverdine, the gene cluster responsible for synthesis of temperature regulated achromobactin siderophore, *acrA* and *acrB*, were also identified in all the three PGPR genomes. In addition to siderophores, chemicals such as phenazine and 4-hydroxybenzoate produced by the PGPRs act as antibiotics and suppress plant pathogenic microbes. In all the three genomes we were able to identify the *phzF* involved in phenazine synthesis and *ubiC* that codes for chorismatelyase involved in 4-hydroxybenzoate synthesis. Also, in CPCRI-2 we identified a homologue of gene associated with the synthesis of anti-microbial compound pyocin [46]. In addition to these, in CPCRI-1 and CPCRI-2 genomes we identified *gabD* and *gabT* involved in production of pest/disease suppressing γ -aminobutyric acid (GABA) [46]. The whole genomes of the three bacteria coded for several genes that encode peroxidases, catalases, superoxide dismutase, and glutathione transferases, all of which alleviate oxidative stress in plants (Fig. S5).

Table 5. List of genes attributable to plant growth promotion traits in the CPCRI PGPR genomes.

Plant growth promotion traits	CPCRI-1	CPCRI-2	CPCRI-3	Genes with potential for conferring PGP traits
Phosphate solubilization	+	+	+	<i>pqq</i> , glucose dehydrogenase gene homolog, <i>pstA</i> , B, C
IAA production	+	–	+	<i>ipdC</i>
Siderophore production	+	+	+	<i>pvd</i> , pyoverdine homologous genes, <i>fpvA</i> , <i>mbtH</i> , <i>acrA</i> , B, <i>fhu</i>
ACC deaminase activity	+	+	+	<i>acdS</i> , <i>rimM</i> , <i>dcyD</i>
Acetoin & butanediol synthesis	+	+	+	<i>als</i> , <i>budA</i> , C, <i>poxB</i>
Phenazine production	+	+	+	<i>phzF</i>
Chitinase production	+	+	+	Chitinase gene homolog
4-hydroxybenzoate production	+	+	+	<i>ubiC</i>
Pyocin	–	+	–	Pyocin gene homolog
Trehalose metabolism	+	+	+	Trehalose synthase gene homolog
H ₂ S production	+	+	+	<i>cysC</i> , J, I, N
Quorum sensing	+	–	+	<i>luxS</i> , <i>Isr</i>
Heat shock proteins	+	+	+	<i>dnaJ</i> , K, <i>groE</i>
Cold shock proteins	+	+	+	<i>cspA</i> , C, D, E
Glycine-betaine production	+	+	+	<i>soxB</i> , <i>opu</i> , <i>proX</i> , glycine betaine gene homolog
Peroxidases	+	+	+	<i>osmC</i> , glutathione peroxidase genes similar to <i>Enterobacterasburiae</i> , <i>oxyR</i>
Catalases	+	+	+	Catalase gene homolog
Superoxide dismutase	+	+	+	<i>sodB</i> , C, super oxide dismutase gene homologs
GABA production	+	+	–	<i>gabD</i> , T

+ indicates presence of genes.

– indicates absence of genes.

doi:10.1371/journal.pone.0104259.t005

Carbohydrate metabolism

Analyses of CPCRI-1, CPCRI-2 and CPCRI-3 genomes showed that they carried genes consistent with their ability to survive in soil environment and plant rhizospheres. The genomes of all 3 bacteria encode genes for central carbohydrate metabolism, including the tricarboxylic acid cycle, the Entner-Doudoroff pathway, glycolysis, gluconeogenesis, pyruvate metabolism and the pentose-phosphate pathways. However, the methyl citrate cycle for propionate metabolism (Table S11) was identified only in CPCRI-2. All the three bacteria carried genes for galactose, fructose, mannose, gluconate and glycogen metabolism, however CPCRI-1 and CPCRI-3 genomes showed the presence of a larger number of metabolic pathways for monosaccharides, disaccharides, oligosaccharides, and polysaccharides, than the CPCRI-2 genome. This indicated that CPCRI-1 and CPCRI-3 could use a large variety of plant-derived carbohydrates as carbon source. Additionally, CPCRI-1 and CPCRI-3 encode genes that can support the use of L-rhamnose, L-arabinose, xylose, trehalose, maltose, lactose and β -glucosides as a carbon source, even though utilization of lactose as a sole carbon source is a characteristic of the *Enterobacteriaceae* family [31]. These findings are consistent with the Biolog studies that demonstrated the ability of CPCRI-1 and CPCRI-3 to use L-rhamnose, trehalose, maltose and lactose.

Trehalose, a disaccharide, is accumulated by many microorganisms growing under high salt or osmotic stress and has been shown to play an important role in *Rhizobium*-legume symbiosis [47]. Accumulation of trehalose in *Bradyrhizobium japonicum* enhances its survival under conditions of salinity stress and plays an important role in the development of symbiotic nitrogen-fixing root nodules on soybean plants [48]. We observed that while all the three PGPRs encoded genes that support trehalose biosyn-

thesis, CPCRI-1 and CPCRI-3 also encoded genes for exogenous trehalose uptake that can potential allow them to use exogenous trehalose.

Malonate metabolism has been characterized in various symbiotic bacteria, such as *A. calcoaceticus*, *K. pneumoniae*, *P. fluorescens* and *P. putida* [49,50]. All three bacterial genomes reported in this study contain genes involved in malonate metabolism. In CPCRI-1 genome, we observed a cluster of nine genes *mdc*ABCDEFGHR (Fig. S6A; CPCRI-1, Gene#3919-3927; CPCRI-3 Gene#2395-2403) involved in malonate decarboxylation (Fig. S6C). Also, CPCRI-2 encodes a nine gene malonate cluster *mdc*MLHGEDCBA (CPCRI-2, Gene#3117-3125; Fig. S6B). This suggests that all the three bacteria are capable of malonate utilization.

Degradation of aromatic compounds

In addition to the carbohydrate metabolism pathway genes, the CPCRI-2 genome coded for genes involved in the degradation of various aromatic compounds such as benzoate, 2,4-dichlorobenzoate, 1,2-dichloroethane, tetrachloroethane and bisphenolA. The β -keto adipate pathway, an important bacterial energy source, has been identified in the *Pseudomonas* species and many members of *Rhizobiaceae* family of soil microorganisms [51,52]. The CPCRI-2 genome contains β -keto adipate pathway genes involved in degradation of lignin derived aromatic compounds. Further in CPCRI-2, we also found genes involved in the metabolism of polyhydroxybutyrate (PHB), an aliphatic polyester synthesized by several bacteria as a means of carbon storage and a source of reducing equivalents in starving conditions [53]. PHB is stored intracellularly as granules and improves bacterial tolerance to high temperatures, H₂O₂ exposure, UV-irradiation, desicca-

tion, and osmotic stress [53,54]. Interestingly, the three genomes also encoded *arsC* gene [45] which may play a role in detoxifying arsenic.

Quorum sensing

Both CPCRI-1 and CPCRI-3 encoded the autoinducer-2 (AI-2) gene [*luxS*; Gene #3755 (CPCRI-1); Gene #3423 (CPCRI-3)]. AI-2 is a small molecule produced by a number of bacterial species, implicated in the regulation of biofilm formation, motility and production of virulence factors [55]. AI-2 has been suggested to act directly through quorum sensing or indirectly through modulation of cellular metabolism. AI-2 dependent quorum sensing system has been demonstrated to be crucial for symbiosis between *Sinorhizobium meliloti* and legumes. *S. meliloti* can respond to the AI-2 signal by up-regulating transcription of its *lsr*-like operon [56]. Genomes of CPCRI-1 and CPCRI-3 also code for *lsr* operon, which contains genes encoding the transport apparatus responsible for internalizing, phosphorylating and processing of the AI-2 signal. The *lsr* operon comprises of six genes, *lsrACDBFG* [57] is present in CPCRI-1 (Gene#2142-2147) and CPCRI-3 (Gene#4093-4098) genomes. The *lsrB* encodes the ligand binding protein, *lsrC* and *lsrD* each encode a transmembrane protein, and *lsrA* encodes a cytoplasmic protein responsible for ATP hydrolysis during transport. In addition to AI-2 quorum systems, non-lux based quorum sensing proteins controlled by *ribB* gene [58] was found in the genomes of all the three CPCRI bacteria (CPCRI-1: Gene#2104, CPCRI-2: Gene#1422, Gene#1433 and for CPCRI-3: Gene#3641). Gene *ribB* is a homolog of the *Escherichia coli* gene for 3,4-dihydroxy-2-butanone 4-phosphate synthase, a key enzyme for riboflavin synthesis, which along with *qsrP*, *acfA*, *qsrV*, and *qsr7* have been proved as non-lux based QSR protein producing genes [58]. In CPCRI-2 (*Pseudomonas putida*) genome, Lux-R system (Gene#3078, 3458, 3799, 3918, 4032, 4115, 4123, 4282, 4600 and 4637) that is involved in acyl-homoserine lactone (ACL) controlled quorum sensing system was found [59]. A collection of genes such as *gacA* (Gene#4600), *rsmA* (Gene#3492) and *rpoS* (Gene#3005) that are known to regulate and network LasRI and RhlRI quorum sensing systems in *Pseudomonas aeruginosa* [60] was also found in CPCRI-2. These observations suggest that CPCRI-1, 2 and 3 may have quorum sensing ability that can contribute to their symbiotic relationship with the host plant.

PGPR fitness conferring genes

Production of heat-shock proteins, cold-shock proteins and osmoregulators in the bacteria regulate survival under harsh conditions. The genomes of all the three CPCRI isolates carried heat-shock protein genes like *dnaJ*, *K* and *groE*, cold-shock proteins genes such as *cspA*, *C*, *D*, and *E*, and several copies of osmoprotectant glycine betaine synthesis genes. Other genes, *gacS* [61], *soxS*, *R*, *oxyR* [62] involved in protecting plants against oxidative stress were also found in the CPCRI genomes. We found *xerC* gene [63] in all the three CPCRI genomes. The *xerC* gene product, a site recombinase, is critical for the PGPRs to be an effective rhizosphere colonizer [63].

Comparison with non-PGPR bacteria

We compared the genes present in CPCRI bacteria with non-PGPR bacteria of similar strain. The CPCRI-1, 3 genomes were compared with *Enterobacter cloacae EcWSU1* [64] and *Enterobacter cloacae subsp. cloacae ATCC 13047* [65], whereas CPCRI-2 genome was compared with *Pseudomonas putida strain S16* [32]. Comparison was performed at functional classification level using GO, SEED and KEGG annotation (Table S12A–I).

Comparison of CPCRI-1, 3 with non-PGPR showed that pyrroloquinoline quinone (*pqq*) biosynthetic gene which is involved in solubilization of mineral phosphates was only present in CPCRI-1, 3 genomes. The acetoin-production gene, which is associated with butanediol dehydrogenase activity, was absent in non-PGPRs. The iron-scavenging group of genes involved in siderophore synthesis and their uptake was more enriched in CPCRI-1, 3 as compared to non-PGPRs. Also, CPCRI-1 genome showed adhesion group of genes to be highly enriched as compared to non-PGPRs.

Comparison of CPCRI-2 with non-PGPR revealed many functional groups that included some key plant growth traits. The widespread colonization island, siderophore enterobactin, pyrroloquinoline quinone (*pqq*) biosynthetic and phenazine (*phz*) biosynthesis genes present in CPCRI-2 were completely absent in *Pseudomonas putida strain S16*. Genes related to adhesion, iron scavenging and sulfur metabolism were more enriched in CPCRI-2 as compared to *Pseudomonas putida strain S16*.

Discussion

In this study we reported the whole genome sequencing and analysis of three PGPRs, CPCRI-1, CPCRI-2 and CPCRI-3 isolated from coconut [16], cocoa [20] and arecanut [22], respectively. The genomic level characterization reported here of PGPRs, to our knowledge is the first for rhizobacteria isolated from coconut, cocoa and arecanut. Usually the bacterial genomes are compact and tightly packed with genes and other functional elements and range in size from 0.5 to 10 Mb, with coding regions averaging ~1 Kb [66]. Following assembly we estimated the genome sizes to be 4.7 Mb for CPCRI-1, 5.9 Mb for CPCRI-2 and 5.1 Mb for CPCRI-3 and there was a good correlation observed between the genome size and genome numbers of the three PGPRs as earlier reported in other studies [26]. The genome size of our *Enterobacter* spp. (CPCRI-1 and 3) was comparable to those of the others isolated from plantation crops such as sugar cane [67] and poplar [31], which had 4.9 and 4.6 Mb sizes respectively. Similarly, the genome size of *Pseudomonas* from cocoa (CPCRI-2) matched with that of the *Pseudomonas aurantiaca* obtained from sugar cane [68]. The GC contents recorded for CPCRI-1 and 3 matched well within the range reported for *Enterobacteriaceae* (38–60%) family and the range reported for the genus *Enterobacter* (52–60%) [69]. Similarly, GC content of CPCRI-2 was observed to fit well in the range expected for *Pseudomonas* genus (58–69%) [70]. Earlier studies have revealed that the GC content of the total genome usually matched with GC content of protein coding genes, spacer genes and stable RNA genes [71]. We could also observe a strong positive correlation between the GC content of the protein coding genes with GC content of total genome of CPCRI-1, 2 and 3 (Table 3, Table S3 and S4). Another interesting observation was about the codon usage pattern: CTG that codes for Leucine (leu) (Fig. S2) was found to be the most preferred codon in the CPCRI isolates as reported earlier in *Escherichia coli* and *Drosophila melanogaster* [72–74].

We identified between 4000 and 4600 protein coding genes in each of the three genomes. While a majority of the genes had homologs in the published sequence database, for 31, 59 and 91 proteins in CPCRI-1, CPCRI-2 and CPCRI-3, respectively, no homologs were found, suggesting that these may have novel functions.

Phylogenetic analysis indicated that both the bacteria isolated from coconut and arecanut belonged to *Enterobacteriaceae* and may reflect the fact that both plantation crops belong to the

Areaceae family and have similar root niche/environment. The cocoa isolate belonged to the *Pseudomonadaceae* family.

Consistent with the PGP properties we found several genes that function in mineral phosphate solubilization, ACC deaminase function, IAA, acetoin and butanediol production. Previously, genes with similar functions in other PGPRs have been reported [36,40,42,75–77]. The genome sequence of *Enterobacter* spp. of coconut and arecanut and *Pseudomonas* from cocoa possessed many genes that have been reported in PGPR isolated from the plantation crops such as poplar and sugarcane. For example, *sodB*, C controlling the superoxide dismutase activity in CPCRI-1 and 3, *oxyR* gene known to regulate production of anti-microbial compound 4-hydroxybenzoate in CPCRI-1, mobility genes *flg*, *flh*, *fim*, and *fli*, in CPCRI-3 had orthologs in *Enterobacter* sp. 638 PGPR isolated from poplar [31]. Similarly, phosphate transporter genes *pstA*, B and C found in CPCRI-1 and 3 had orthologs in the *Enterobacter* spp. SP1 PGPR isolated from sugarcane [67]. Additionally, comparison of CPCRI-1, 2 and 3 genomes against non-PGPR genomes of the same genus showed several plant growth related group of genes that were either absent, like the pyrroloquinoline quinone (*pqq*) biosynthetic process gene, or less enriched in non-PGPR genomes.

In addition to growth promoting functions, PGPRs also indirectly support plant growth by suppressing pathogens [2]. In the PGPR genomes reported in this study, we identified several genes that are known to support the production of antimicrobial compounds such as siderophores, phenazine, 4-hydroxybenzoate and GABA [46]. They also contained genes for chitinase enzyme that can dissolve cell walls of pathogenic fungi, nematodes and insect pests [46]. In addition, CPCRI-2 genome encoded a gene for production of pyocin, a compound that suppresses growth of other related species. The three PGPR genomes also encoded enzymes such as peroxidases, catalases, super oxide dismutases and glutathione transferases all of which are involved in the management of oxidative stresses in plants.

Sulfur is an essential nutrient for plant growth and development and is associated with stress tolerance in plants [78]. Crop plants generally rely on the soil for their sulfur requirement and the mobilization of this sulfur for assimilation by plants is mediated by the microbial community in the soil and rhizosphere [79]. Sulfur-deficient conditions can cause severe losses in crop yield [80]. Sulfur nutrition is demonstrated to be critical in cocoa somatic embryogenesis [81]. In cocoa, elemental sulfur was identified in the xylem of resistant genotypes after infection by the vascular fungal pathogen *Verticillium dahlia* [82]. We found genes involved in H₂S biosynthesis in all the three PGPRs sequenced and they may, in particular in cocoa PGPR (CPCRI-2), be an important source of sulfur. We have also identified protein coding genes in the three bacteria known to be involved in resistance to copper, cobalt, zinc, arsenic, mercury and cadmium, suggesting that they function in detoxification of these metals.

Sequence analysis also showed that all the three CPCRI isolates have complete gene clusters corresponding to Type II, VI, Sec and Twin arginine targeting gene complexes (Table S13). Some of the past studies have shown that the Type I–VI and Sec secretion systems in rhizobacteria *Pseudomonas fluorescens* and *Variovorax paradoxus* function in promoting plant growth [45,46,83,84]. The presence of these secretion systems in PGPRs may play a role in their plant growth promoting functions and also provide support for their rhizosphere colonization ability [85,86].

Among the many biological properties of CPCRI isolates, their ability to utilize different carbohydrate sources and survive and grow under a wide range of pH, NaCl concentrations, and temperature would be able to help them establish well under

changing soil conditions. Accumulation of disaccharide trehalose has been implicated in survival of some of the plant-beneficial symbiotic microorganisms under salt or osmotic stress conditions [47,48]. We observed that while all the three bacteria are capable of trehalose biosynthesis, CPCRI-1 and CPCRI-3 also have genes (*treY*, *Z*) that will support exogenous trehalose uptake, further indicating that they are capable of tolerating high salinity or osmotic stress. Presence of genes that regulate the production of heat-shock, cold-shock proteins and osmoregulants in CPCRI PGPRs indicate that they have the capabilities to adapt to harsh conditions for their survival.

The genomic information obtained support the observed traits making them ideal candidates for further development as biofertilizers. The genes identified in our draft genome can now be studied for specific functions using knockout strategies. Experiments can be designed to identify the genes involved in the plant colonization and plant growth promotion process. Genetic engineering can be used to further improve the plant growth promoting properties of these bacteria. These findings will help in designing comprehensive strategies for development and use of such PGPRs to support sustainable plantation crop cultivation.

Materials and Methods

PGPR strains

About 1512 morphologically distinct heterotrophic bacteria were isolated from coconut, cocoa and arecanut rhizosphere soil samples [13,15,18,22] collected from privately owned farms with the permission of the owner. The different agro-ecological zones in India from which the samples were collected are listed in Table S1. The bacteria were first screened *in vitro* for a dozen important plant growth promoting properties and then tested for growth promotion in rice (for coconut and arecanut isolates) and cowpea (for cocoa isolates). Also, they were tested for growth promotion activity in coconut, cocoa and arecanut seedlings [13,16,18,19]. Based on their plant growth promotion characteristics, three PGPRs, designated here as CPCRI-1 (from coconut), CPCRI-2 (from cocoa) and CPCRI-3 (from arecanut) were given bio labels as RNF267 [16], KGSF20 [20], and KtRA5-88 [22], respectively, based on place/source of isolation and were selected for whole genome sequencing studies.

Bacterial cell morphology was assessed microscopically. Gram's staining was also performed. PGPR identification was done by conventional biochemical assays and Biolog analysis [18,19]. Cultures grown for 24 h on Biolog universal growth (BUG) agar were collected and processed according to the manufacturer's instructions (Hayward, CA). Briefly, cultures were transferred to inoculating fluid A (IF-A) and inoculum density was adjusted to 98% T using Biolog turbidimeter (Hayward, CA). Using multi-channel pipette, cell suspension was inoculated into Biolog Gen III Microplates (100 µl/well) containing 96 wells that provides 94 phenotypic tests. Plates were incubated at 33°C for 24 h. The optical density at 590 nm produced from the reduction of tetrazolium violet in each well was read after 24 h using a Biolog Microplate reader (version 5.1.1). Identification was performed by comparing the pattern formed in culture wells with possible patterns in the Microstation/MicrologVersion 5.1.1 database. A species identification of the PGPRs isolated coconut, cocoa and arecanut was acknowledged when the similarity index (SIM) and distance (DIS) values were >0.5 and <5.0, respectively [13,15,16,19,20].

Genomic DNA isolation

The three PGPRs, CPCRI-1, CPCRI-2 and CPCRI-3, chosen based on their plant beneficial attributes towards coconut, cocoa and arecanut were grown in Tryptic Soy Broth (TSB) medium at 30°C for 24–48 h. Genomic DNA was extracted using Gen Elute bacterial genomic DNA kit (Sigma, USA) as per the manufacturer's instructions. The extracted DNA was resolved on 0.8% agarose gel to check its integrity. The quality of the genomic DNA samples was assessed using Bioanalyzer DNA 7500 chip (Agilent, CA). The DNA yield was estimated on a TBS-380 Mini-Fluorometer (Turner BioSystems, CA) using PicoGreen dsDNA Quantitation Reagent (Molecular Probes, OR)

Library preparation and multiplexed whole genome shotgun sequencing

Whole genome shotgun libraries were generated from 1 µg genomic DNA using the GS FLX Titanium Rapid Library Preparation Kit (Roche Applied Science, CA) according to the manufacturer's protocol. Rapid library MID Adaptors MID10, MID11 and MID12 (Roche Applied Science, CA) were ligated to the CPCRI-1, CPCRI-2 and CPCRI-3 libraries, respectively. The quality of the libraries (library size ~1,600 bp) was assessed using Bioanalyzer High sensitivity DNA chip (Agilent, CA). The libraries were titrated by emulsion titrations and based on the percent enrichment, appropriate amount of the libraries were used to set up the large volume emulsion PCRs for each of the individual libraries. The beads containing clonally amplified DNA were enriched and the sequencing primer was annealed. Finally, half of the beads containing the CPCRI-2 libraries were mixed with CPCRI-1 library beads and the other half with the CPCRI-3 library beads. Each set of bead mix was then loaded on a picoTiter plate (half the plate) and sequenced using the GS FLX Titanium Sequencing Kit XL+ (Roche Applied Science, CA). Upon sequencing and processing of the raw data, demultiplexed data were assembled using GS de novo assembler version 2.6 (Roche Applied Science, CA).

Gene prediction

The protein-coding genes prediction was performed using Glimmer-MG [25], a metagenomics gene prediction program that uses interpolated Markov models (IMMs) to identify the protein-coding regions in the genome. The default setting for Glimmer-MG was used for gene prediction. The tRNA genes in the genome were identified using tRNA-SE program [87]. The BLASTN program (E-value $= 1.0 \times 10^{-10}$) at WebMGA was used for predicting ribosomal RNA genes [88].

Gene annotation

For gene annotation, we first compared predicted protein-coding genes against the non-redundant (nr) NCBI protein database using BLASTX (E-value $= 1.0 \times 10^{-5}$) program [27,89]. BLASTX result was parsed and the top hit database accession numbers were extracted. The accession numbers were then compared against UniProt knowledgebase for annotating genes [90]. The BLASTX result was imported into MEGAN4 [33,91] to perform KEGG pathway analysis [92] and SEED classification [93] of the proteins. The annotated genes were inspected for identifying those involved in PGP functions, pathogen suppression, abiotic stress tolerance, rhizosphere competence, carbohydrate metabolism and other important relevant functions.

Phylogenetic analysis

Phylogenetic analysis was performed using AMPHORA2 [24,94], a phylogenomic inference tool used for genomic

phylotyping of bacteria and archaeal genomes. It scans the genome for 31 marker genes, which are universally distributed in both phyla. The 31 marker genes identified in CPCRI-1, CPCRI-2, and CPCRI-3 genome were then aligned using ClustalW [95]. Phylogenetic tree was inferred using bootstrap method available in ClustalW package.

Genome comparison

We compared our assembled bacterial genomes with the available complete bacterial genomes using progressive Mauve aligner [34] using default settings. The published genomes used in the alignment were obtained from PATRIC database (<http://patricbrc.vbi.vt.edu>) [96]. The sequence alignment file generated by the aligner was parsed to calculate pairwise similarity. Briefly, we first extracted the conserved blocks from the alignment file and then regions with <math>< 50</math> continuous gaps were considered for computing similarity score based on a pairwise sequence similarity percentage and coverage score which represents the percentage of genome that could be aligned pairwise.

Supporting Information

Figure S1 CPCRI genomes. Circos plot representing the CPCRI-1 (A), CPCRI-2 (B) and CPCRI-3 (C) genomes. The innermost circle represents the GC content, the second circle from the innermost circle represent non-coding genes, the third circle from inside represents coding genes on negative strand, the fourth circle represents coding genes on positive strand, and the outermost circle represent contigs.

(TIF)

Figure S2 GC-content based on codon position. GC-content distribution at each of the three codon position derived from proportion of genes with a given GC-content at that position is shown for CPCRI-1 (A) CPCRI-2 (B) and CPCRI-3 (C).

(TIF)

Figure S3 Codon usage. The proportion of each codon (%) used in the CPCRI PGPR genomes computed from the protein-coding genes.

(TIF)

Figure S4 Protein taxonomy tree. Proteins encoded by the CPCRI PGPR genomes analyzed using MEGAN4 [33]. The numbers in bracket represent total number of gene assigned based on MEGAN4 annotation. The number in the bracket correspond to CPCRI-1, CPCRI-2 and CPCRI-3 in that order.

(TIFF)

Figure S5 Number of genes coding for oxidative stress response enzymes in each of the indicated CPCRI PGPR strains.

(TIF)

Figure S6 Malonate gene cluster in (A) CPCRI-1, (B) CPCRI-2, (C) CPCRI-3 genome.

(TIFF)

Table S1 Isolation details of PGPR isolates.

(XLSX)

Table S2 Biological properties of PGPRs (CPCRI-1, CPCRI-2, CPCRI-3).

(XLSX)

Table S3 Genomic properties of *Enterobacteriaceae* family.

(XLSX)

Table S4 Genomic properties of *Pseudomonadeceae* family.

(XLSX)

Table S5 **A.** Predicted protein-coding genes in CPCRI-1. **B.** Predicted protein-coding genes in CPCRI-2. **C.** Predicted protein-coding genes in CPCRI-3. (XLSX)

Table S6 **A.** Predicted tRNA genes in CPCRI-1 genome. **B.** Predicted tRNA genes in CPCRI-2 genome. **C.** Predicted tRNA genes in CPCRI-3 genome. (XLSX)

Table S7 Predicted rRNA genes in CPCRI-1, CPCRI-2, CPCRI-3 genome. (XLSX)

Table S8 **A.** CPCRI-1 protein coding genes annotation. **B.** CPCRI-2 protein coding genes annotation. **C.** CPCRI-3 protein coding genes annotation. (XLSX)

Table S9 Pairwise comparison of CPCRI-1, CPCRI-2, and CPCRI-3 PGPR genomes against bacteria genomes using progressive Mauve aligner. Percentage sequence similarity/genome coverage in the conserved block is shown. (XLSX)

Table S10 List of GO terms identified in the bacterial genomes. (XLSX)

Table S11 Propionate metabolism pathway genes in CPCRI-1, CPCRI-2 and CPCRI-3 PGPR genomes. (XLSX)

Table S12 **A.** SEED comparison summary for CPCRI-1, 3 vs non-PGPRs (*Enterobacter cloacae EcWSU1*, *Enterobacter cloacae subsp. cloacae ATCC 13047*). **B.** KEGG comparison summary for CPCRI-1, 3 vs non-PGPRs (*Enterobacter cloacae EcWSU1*,

Enterobacter cloacae subsp. cloacae ATCC 13047). **C.** SEED comparison summary for CPCRI-2 vs non-PGPRs (*Pseudomonas putida S16*). **D.** KEGG comparison summary for CPCRI-2 vs non-PGPRs (*Pseudomonas putida S16*) **E.** GO comparison summary of CPCRI-1 vs *Enterobacter cloacae EcWSU1* (a non-PGPR). **F.** GO comparison summary of CPCRI-1 vs *Enterobacter cloacae subsp. cloacae ATCC 13047* (a non-PGPR). **G.** GO comparison summary of CPCRI-2 vs *Pseudomonas putida strain S16* (a non-PGPR). **H.** GO comparison summary of CPCRI-3 vs *Enterobacter cloacae EcWSU1* (a non-PGPR). **I.** GO comparison summary of CPCRI-3 vs *Enterobacter cloacae subsp. cloacae ATCC 13047* (a non-PGPR). (XLSX)

Table S13 Bacteria secretion system KEGG pathway gene in CPCRI-1, CPCRI-2 and CPCRI-3 genomes. (XLSX)

Result S1 Protein taxonomy results using MEGAN4 program. (DOCX)

Acknowledgments

We thank Sam Santhosh for the support and being the driving force behind the project. We are grateful to Devi Santhosh and Sneha Somasekar for helping edit the manuscript.

Author Contributions

Conceived and designed the experiments: AG MG PR RG. Wrote the paper: GT AG MG GVT SS PR RG. Isolated the genomic DNA from selected PGPR strains: AG MG. Performed the sequencing: JG PR. Performed the analysis: MG VM RG. Provided technical advice, analysis support and oversight: SS GVT SCS.

References

- Kloepper JW, Schroth MN (1978) Plant growth-promoting rhizobacteria on radishes. 4th Internat Conf on Plant Pathogenic Bacter Station de Pathologie Vegetale et Phytobacteriologie, INRA, Angers, France. pp. 879–882
- Lugtenberg B, Kamilova F (2009) Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* 63: 541–556.
- Nelson LM (2004) Plant growth promoting rhizobacteria (PGPR): Prospects for new inoculants. *Crop Management* 3: 301–305.
- Bloemberg GV, Lugtenberg BJ (2001) Molecular basis of plant growth promotion and biocontrol by rhizobacteria. *Curr Opin Plant Biol* 4: 343–350.
- Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nat Methods* 5: 16–18.
- MacLean D, Jones JD, Studholme DJ (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* 7: 287–296.
- Mathimaran N, Srivastava R, Wiemken A, Sharma AK, Boller T (2012) Genome sequences of two plant growth-promoting fluorescent *Pseudomonas* strains, R62 and R81. *J Bacteriol* 194: 3272–3273.
- Song JY, Kim HA, Kim JS, Kim SY, Jeong H, et al. (2012) Genome Sequence of the Plant Growth-Promoting Rhizobacterium *Bacillus* sp. Strain JS. *J Bacteriol* 194: 3760–3761.
- Ma M, Wang C, Ding Y, Li L, Shen D, et al. (2011) Complete genome sequence of *Paenibacillus polymyxa* SC2, a strain of plant growth-promoting Rhizobacterium with broad-spectrum antimicrobial activity. *J Bacteriol* 193: 311–312.
- Matilla MA, Pizarro-Tobias P, Roca A, Fernandez M, Duque E, et al. (2011) Complete genome of the plant growth-promoting rhizobacterium *Pseudomonas putida* BIRD-1. *J Bacteriol* 193: 1290.
- Bopaiiah BM (1985) Occurrence of phosphate solubilising microorganisms in the root region of arecanut palms. *J Plantation Crops* 13: 60–62.
- Bopaiiah BM, Shetty HS (1991) Soil microflora and biological activities in the rhizospheres and root regions of coconut-based multistoreyed cropping and coconut monocropping systems. *Soil Biol Biochem* 23: 89–94.
- George P, Gupta A, Gopal M, Thomas L, Thomas GV (2012) Screening and in vitro evaluation of phosphate solubilizing bacteria from rhizosphere and roots of coconut palms (*Cocos nucifera* L.) growing in different states of India. *Journal of Plantation Crops* 40: 61–65.
- Ghai SK, Thomas GV (1989) Occurrence of Azospirillum in coconut based farming systems. *Plant and Soil* 114: 235–241.
- Thomas L, Gupta A, Gopal M, ChandraMohan R, George P, et al. (2011) Evaluation of rhizospheric and endophytic *Bacillus* spp. and fluorescent *Pseudomonas* spp. isolated from *Theobroma cacao* L. for antagonistic reaction to *Phytophthora palmivora*, the causal organism of black pod disease of cocoa. *J Plantation Crops* 39: 370–376.
- George P, Gupta A, Gopal M, Thomas L, Thomas GV (2013) Multifarious beneficial traits and plant growth promoting potential of *Serratia marcescens* KiSII and *Enterobacter* sp. RNF 267 isolated from the rhizosphere of coconut palms (*Cocos nucifera* L.). *World J Microbiol Biotechnol* 29: 109–117.
- Gupta A, Gopal M, Thomas GV (2006) Bioaugmentation of *Cocos nucifera* L. seedlings with the plant growth promoting rhizobacteria, *Bacillus coagulans* and *Brevibacillus brevis* for growth promotion. XXXVI Conference and Annual Meeting of European Society for New Methods in Agricultural Research. Iasi, Romania.
- Thomas L, Gopal M, Thomas GV (2013) Consolidated report (2009–2013) of the NAIP Project on Diversity analysis of *Bacillus* and other predominant genera in extreme environments and their utilization in agriculture. Central Plantation Crops Research Institute, Kasaragod, Kerala, India.
- Thomas GV, Gupta A, Gopal M (2014) Consolidated report (2006–2014) of the ICAR Network Project on Application of Microorganisms in Agriculture and Allied Sectors: Development and application of PGPR formulations for growth improvement and disease suppression in coconut and cocoa. Central Plantation Crops Research Institute, Kasaragod, Kerala, India.
- Thomas L (2013) Identification and evaluation of plant growth promoting rhizobacteria from the diverse bacilli and fluorescent pseudomonad population in rhizosphere and roots of cocoa (*Theobroma cacao* L.). PhD Thesis, Mangalore University, Karnataka, India.
- Thomas GV, Iyer R, Bopaiiah BM (1991) Beneficial microbes in the nutrition of coconut. *J Plantation Crops* 19: 127–138.
- Anusree GK, Gupta A, Gopal M, Manjusha A, Thomas GV. Isolation and characterization of acid tolerant bacteria from the rhizosphere of arecanut growing in extremely acidic soils of Kerala and Karnataka. In preparation.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Wu M, Scott AJ (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28: 1033–1034.
- Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL (2012) Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res* 40: e9.

26. Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci U S A* 101: 3160–3165.
27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
28. Jones P, Binns D, Chang HY, Fraser M, Li W, et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*.
29. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39: D225–229.
30. UniProt C (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42: D191–198.
31. Taghavi S, van der Lelie D, Hoffman A, Zhang YB, Walla MD, et al. (2010) Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp. 638. *PLoS Genet* 6: e1000943.
32. Yu H, Tang H, Wang L, Yao Y, Wu G, et al. (2011) Complete genome sequence of the nicotine-degrading *Pseudomonas putida* strain S16. *J Bacteriol* 193: 5541–5542.
33. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* 21: 1552–1560.
34. Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5: e11147.
35. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33: 5691–5702.
36. Rodriguez H, Fraga R, Gozalez T, Bashan Y (2006) Genetics of phosphate solubilization and its potential applications for improving plant growth-promoting bacteria. *Plant Soil* pp. 15–21.
37. Duca D, Lörj J, Patten CL, Rose D, Glick BR (2014) Indole-3-acetic acid in plant-microbe interactions. *Antonie Van Leeuwenhoek* 106: 85–125.
38. Straub D, Yang H, Liu Y, Tsap T, Ludewig U (2013) Root ethylene signalling is involved in *Miscanthus sinensis* growth promotion by the bacterial endophyte *Herbaspirillum frisingense* GSF30(T). *J Exp Bot* 64: 4603–4615.
39. Shah S, Li J, Moffatt BA, Glick BR (1998) Isolation and characterization of ACC deaminase genes from two different plant growth-promoting rhizobacteria. *Can J Microbiol* 44: 833–843.
40. Medigue C, Krin E, Pascal G, Barbe V, Bernsel A, et al. (2005) Coping with cold: the genome of the versatile marine Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 15: 1325–1335.
41. Dooley FD, Nair SP, Ward PD (2013) Increased growth and germination success in plants following hydrogen sulfide administration. *PLoS One* 8: e62048.
42. Ryu CM, Farag MA, Hu CH, Reddy MS, Wei HX, et al. (2003) Bacterial volatiles promote growth in *Arabidopsis*. *Proc Natl Acad Sci U S A* 100: 4927–4932.
43. Blomqvist K, Nikkola M, Lehtovaara P, Suihko ML, Airaksinen U, et al. (1993) Characterization of the genes of the 2,3-butanediol operons from *Klebsiella terrigena* and *Enterobacter aerogenes*. *J Bacteriol* 175: 1392–1404.
44. Renna MC, Najmudin N, Winik LR, Zahler SA (1993) Regulation of the *Bacillus subtilis* *alsS*, *alsD*, and *alsR* genes involved in post-exponential-phase production of acetoin. *J Bacteriol* 175: 3863–3875.
45. Duan J, Jiang W, Cheng Z, Heikkilä JJ, Glick BR (2013) The complete genome sequence of the plant growth-promoting bacterium *Pseudomonas* sp. UW4. *PLoS One* 8: e58640.
46. Loper JE, Hassan KA, Mavrodi DV, Davis EW 2nd, Lim CK, et al. (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* 8: e1002784.
47. Zahran HH (1999) Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiol Mol Biol Rev* 63: 968–989, table of contents.
48. Sugawara M, Cytryn EJ, Sadowsky MJ (2010) Functional role of *Bradyrhizobium japonicum* trehalose biosynthesis and metabolism genes during physiological stress and nodulation. *Appl Environ Microbiol* 76: 1071–1081.
49. Kim YS (2002) Malonate metabolism: biochemistry, molecular biology, physiology, and industrial application. *J Biochem Mol Biol* 35: 443–451.
50. Koo JH, Kim YS (1999) Functional evaluation of the genes involved in malonate decarboxylation by *Acinetobacter calcoaceticus*. *Eur J Biochem* 266: 683–690.
51. Li D, Yan Y, Ping S, Chen M, Zhang W, et al. (2010) Genome-wide investigation and functional characterization of the beta-ketoadipate pathway in the nitrogen-fixing and root-associated bacterium *Pseudomonas stutzeri* A1501. *BMC Microbiol* 10: 36.
52. MacLean AM, MacPherson G, Aneja P, Finan TM (2006) Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. *Appl Environ Microbiol* 72: 5403–5413.
53. Kadowaki MA, Muller-Santos M, Rego FG, Souza EM, Yates MG, et al. (2011) Identification and characterization of PhbF: a DNA binding protein with regulatory role in the PHB metabolism of *Herbaspirillum seropedicae* SmR1. *BMC Microbiol* 11: 230.
54. Ratcliff WC, Kadam SV, Denison RF (2008) Poly-3-hydroxybutyrate (PHB) supports survival and reproduction in starving rhizobia. *FEMS Microbiol Ecol* 65: 391–399.
55. Reading NC, Sperandio V (2006) Quorum sensing: the many languages of bacteria. *FEMS Microbiol Lett* 254: 1–11.
56. Pereira CS, McAuley JR, Taga ME, Xavier KB, Miller ST (2008) *Sinorhizobium meliloti*, a bacterium lacking the autoinducer-2 (AI-2) synthase, responds to AI-2 supplied by other bacteria. *Mol Microbiol* 70: 1223–1235.
57. Pereira CS, de Regt AK, Brito PH, Miller ST, Xavier KB (2009) Identification of functional LsrB-like autoinducer-2 receptors. *J Bacteriol* 191: 6975–6987.
58. Callahan SM, Dunlap PV (2000) LuxR- and acyl-homoserine-lactone-controlled non-lux genes define a quorum-sensing regulon in *Vibrio fischeri*. *J Bacteriol* 182: 2811–2822.
59. Fuqua C, Winans SC, Greenberg EP (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu Rev Microbiol* 50: 727–751.
60. Schuster M, Greenberg EP (2006) A network of networks: quorum-sensing gene regulation in *Pseudomonas aeruginosa*. *Int J Med Microbiol* 296: 73–81.
61. Whistler CA, Corbell NA, Sarniguet A, Ream W, Loper JE (1998) The two-component regulators GacS and GacA influence accumulation of the stationary-phase sigma factor sigmaS and the stress response in *Pseudomonas fluorescens* Pf-5. *J Bacteriol* 180: 6635–6641.
62. Ochsner UA, Vasil ML, Alsabbagh E, Parvatiyar K, Hassett DJ (2000) Role of the *Pseudomonas aeruginosa* oxyR-recG operon in oxidative stress defense and DNA repair: OxyR-dependent regulation of *katB*-*ankB*, *ahpB*, and *ahpC*-*ahpF*. *J Bacteriol* 182: 4533–4544.
63. Shen X, Hu H, Peng H, Wang W, Zhang X (2013) Comparative genomic analysis of four representative plant growth-promoting rhizobacteria in *Pseudomonas*. *BMC Genomics* 14: 271.
64. Humann JL, Wildung M, Cheng CH, Lee T, Stewart JE, et al. (2011) Complete genome of the onion pathogen *Enterobacter cloacae* EcWSU1. *Stand Genomic Sci* 5: 279–286.
65. Ren Y, Ren Y, Zhou Z, Guo X, Li Y, et al. (2010) Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* type strain ATCC 13047. *J Bacteriol* 192: 2463–2464.
66. Ochman H, Davalos LM (2006) The nature and dynamics of bacterial genomes. *Science* 311: 1730–1733.
67. Zhu B, Chen M, Lin L, Yang L, Li Y, et al. (2012) Genome sequence of *Enterobacter* sp. strain SP1, an endophytic nitrogen-fixing bacterium isolated from sugarcane. *J Bacteriol* 194: 6963–6964.
68. Mehraz S, Bauer JS, Gross H (2014) Complete Genome Sequence of the Sugar Cane Endophyte *Pseudomonas aurantiaca* PB-St2, a Disease-Suppressive Bacterium with Antifungal Activity toward the Plant Pathogen *Colletotrichum falcatum*. *Genome Announc* 2.
69. Rameshkumar N, Lang E, Nair S (2010) Mangrovibacter plantisponsor gen. nov., sp. nov., a nitrogen-fixing bacterium isolated from a mangrove-associated wild rice (*Porteresia coarctata* Tateoka). *Int J Syst Evol Microbiol* 60: 179–186.
70. Couillerot O, Prigent-Combaret C, Caballero-Mellado J, Moenne-Loccoz Y (2009) *Pseudomonas fluorescens* and closely-related fluorescent pseudomonads as biocontrol agents of soil-borne phytopathogens. *Lett Appl Microbiol* 48: 505–512.
71. Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A* 84: 166–169.
72. Hershberg R, Petrov DA (2009) General rules for optimal codon choice. *PLoS Genet* 5: e1000556.
73. Vicario S, Moriyama EN, Powell JR (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol Biol* 7: 226.
74. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, et al. (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res* 16: 8207–8211.
75. Shankar M, Ponraj P, Ilakiam D, Rajendran J, Gunasekaran P (2012) Genome sequence of the plant growth-promoting bacterium *Enterobacter cloacae* GS1. *J Bacteriol* 194: 4479.
76. Sharma V, Kumar V, Archana G, Kumar GN (2005) Substrate specificity of glucose dehydrogenase (GDH) of *Enterobacter asburiae* PSI3 and rock phosphate solubilization with GDH substrates as C sources. *Can J Microbiol* 51: 477–482.
77. Goldstein AH (1996) Involvement of the quinoprotein glucose dehydrogenase in the solubilization of exogenous phosphates by Gram-negative bacteria; Washington, DC. ASM Press. pp. 197–203.
78. Gill SS, Tuteja N (2011) Cadmium stress tolerance in crop plants: probing the role of sulfur. *Plant Signal Behav* 6: 215–222.
79. Kertesz MA, Fellows E, Schmalenberger A (2007) Rhizobacteria and plant sulfur supply. *Adv Appl Microbiol* 62: 235–268.
80. Nemat MA, El-Kader AAA, Attia M, Alva AK (2011) Effects of Nitrogen Fertilization and Soil Inoculation of Sulfur-Oxidizing or Nitrogen-Fixing Bacteria on Onion Plant Growth and Yield. *International Journal of Agronomy* 2011.
81. Emile M, Nicolas N, Auguste IE, Abdourahmane S, Omokolo DN (2010) Sulphur depletion altered somatic embryogenesis in *Theobroma cacao* L. Biochemical difference related to sulphur metabolism between embryogenic and non embryogenic calli. *African Journal of Biotechnology* 9: 5665–5675.
82. Cooper RM, Resende MLV, Flood J, Rowan MG, Beale MH, et al. (1996) Detection and cellular localization of elemental sulphur in disease-resistant genotypes of *Theobroma cacao*. *Nature* 379: 159–162.
83. Preston GM, Bertrand N, Rainey PB (2001) Type III secretion in plant growth-promoting *Pseudomonas fluorescens* SBW25. *Mol Microbiol* 41: 999–1014.

84. Han JI, Choi HK, Lee SW, Orwin PM, Kim J, et al. (2011) Complete genome sequence of the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. *J Bacteriol* 193: 1183–1190.
85. Viollet A, Corberand T, Mougel C, Robin A, Lemanceau P, et al. (2011) Fluorescent pseudomonads harboring type III secretion genes are enriched in the mycorrhizosphere of *Medicago truncatula*. *FEMS Microbiol Ecol* 75: 457–467.
86. Barret M, Egan F, Moynihan J, Morrissey JP, Lesouhaitier O, et al. (2013) Characterization of the SPI-1 and Rsp type three secretion systems in *Pseudomonas fluorescens* F113. *Environ Microbiol Rep* 5: 377–386.
87. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33: W686–689.
88. Wu S, Zhu Z, Fu L, Niu B, Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12: 444.
89. Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3: 266–272.
90. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009.
91. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17: 377–386.
92. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–114.
93. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75.
94. Wu M, Eisen JA (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* 9: R151.
95. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
96. Snyder EE, Kampanya N, Lu J, Nordberg EK, Karur HR, et al. (2007) PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res* 35: D401–406.