SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
SOIL MICROBIOLOGY
PHYLOGENETICS

Received
9 June 2014

Accepted
11 August 2014

Published
2 September 2014

Correspondence and requests for materials should be addressed to Y.M.-L. (yvan.moenne-loccoz@univ-lyon1.fr)

Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria

Maxime Bruto[1,2,3], Claire Prigent-Combaret[1,2,3], Daniel Muller[1,2,3] & Yvan Moënne-Loccoz[1,2,3]

[1]Université de Lyon, F-69622, Lyon, France, [2]Université Lyon 1, Villeurbanne, France, [3]CNRS, UMR5557, Ecologie Microbienne, Villeurbanne, France.

The positive effects of root-colonizing bacteria cooperating with plants lead to improved growth and/or health of their eukaryotic hosts. Some of these Plant Growth-Promoting Rhizobacteria (PGPR) display several plant-beneficial properties, suggesting that the accumulation of the corresponding genes could have been selected in these bacteria. Here, this issue was targeted using 23 genes contributing directly or indirectly to established PGPR effects, based on genome sequence analysis of 304 contrasted Alpha- Beta- and Gammaproteobacteria. Most of the 23 genes studied were also found in non-PGPR Proteobacteria and none of them were common to all 25 PGPR genomes studied. However, ancestral character reconstruction indicated that gene transfers -predominantly ancient- resulted in characteristic gene combinations according to taxonomic subgroups of PGPR strains. This suggests that the PGPR-plant cooperation could have established separately in various taxa, yielding PGPR strains that use different gene assortments. The number of genes contributing to plant-beneficial functions increased along the continuum -animal pathogens, phytopathogens, saprophytes, endophytes/symbionts, PGPR- indicating that the accumulation of these genes (and possibly of different plant-beneficial traits) might be an intrinsic PGPR feature. This work uncovered preferential associations occurring between certain genes contributing to phytobeneficial traits and provides new insights into the emergence of PGPR bacteria.

Plant roots host a large variety of bacteria, many of them cooperating with the plant and enhancing plant nutrition, stress tolerance or health[1]. Several different modes of action are documented in these Plant Growth-Promoting Rhizobacteria (PGPR). Direct effects on plants may involve enhanced availability of nutriments[2,3], stimulation of root system development via production of phytohormones and other signals[4] or interference with plant's ethylene synthesis[5,6], and/or induced systemic resistance[7]. Indirect beneficial effects of PGPR on plants entail competition or antagonism towards phytoparasites[8,9].

Despite extensive literature on PGPR's modes of action (especially in the Proteobacteria), the molecular features that define a PGPR remain elusive, because the PGPR status is not always well defined. First, PGPR may occupy different microbial habitats, as they range from saprophytic soil bacteria that colonize the rhizosphere to bacteria that can also colonize internal root tissues. This means that the distinction is not often simple respectively with saprophytes without plant-beneficial effects (especially plant commensals) and with vertically-inherited endophytes or plant endosymbionts. Second, several bacteria display alternate ecological niches, and at times some may function as PGPR. For instance, certain tumor-inducing *Agrobacterium* strains have plant growth stimulation potential on non-susceptible plant hosts[10], a property also found in an *Escherichia coli* gut commensal[10]. Third, the genes implicated in plant-beneficial functions range from genes directly conferring plant-beneficial properties, such as *nif* (nitrogen fixation)[11] or *phl* (phloroglucinol synthesis)[12], to genes contributing to a variety of cell functions indirectly or secondarily including plant-beneficial ones, such as *pqq* (pyrroloquinoline quinone synthesis)[13]. Fourth, many PGPR strains are not yet recognized as such (as determination of PGPR status requires experimental assessment), and it is very likely that not all plant-beneficial traits and the corresponding genes have already been identified. Fifth, the assessment of genes encoding plant-beneficial properties is commonly restrained to particular bacterial clades[14] if not particular PGPR strains[9,12], without a more general analysis of gene distribution across several bacterial clades[15].

Table 1 | Distribution of plant-beneficial function contributing (PBFC) genes according to the primary ecological lifestyle documented for the bacteria studied

Gene function	*Gene*	*PGPR (25[a])*	*Endophytes/symbionts (56)*	*Saprophytes (29)*	*Phytopathogens (59)*	*Animal pathogens (135)*
Phosphate solubilization	*pqqB*	20	36	26	34	13
	pqqC	20	36	26	35	13
	pqqD	20	36	26	35	13
	pqqE	20	36	26	35	13
	pqqF	10	17	8	16	7
	pqqG	7	17	9	13	4
2,4-Diacetylphloroglucinol synthesis	*phlA*	3	0	0	0	0
	phlB	3	0	0	0	0
	phlC	3	0	0	0	0
	phlD	3	0	0	0	4
Hydrogen cyanide synthesis	*hcnA*	3	9	2	0	4
	hcnB	3	9	2	0	4
	hcnC	3	9	2	0	4
Acetoine/2,3-butanediol synthesis	*budA*	5	2	3	14	5
	budB	5	2	3	14	5
	budC	11	12	4	10	5
Nitric oxide synthesis	*nirK*	6	14	1	1	108
Auxin synthesis	*ipdC*	5	2	3	10	5
	ppdC	2	2	0	0	0
ACC deamination	*acdS*	9	31	16	26	44
Nitrogen fixation	*nifD*	9	23	3	3	0
	nifH	9	23	3	3	0
	nifK	9	23	3	3	0

[a]The number of bacteria is indicated in parenthesis.

Despite these limitations, however, a number of emblematic PGPR model strains have been extensively characterized over the last 20 years, uncovering the molecular basis of at least some of their plant-beneficial effects. These studies have evidenced that many PGPR strains typically harbor more than one plant-beneficial property[8,16], and it could be hypothesized that the accumulation of genes contributing (whether directly or indirectly) to plant-beneficial traits has been selected by the interaction of these bacteria with plants. On this basis, it could even be expected that PGPR might be identified by their particular assortment of genes contributing to plant-beneficial functions. So far, a more general description of the occurrence of these genes, including in bacteria not interacting with plants, is still lacking. Such knowledge would bring fundamental insights into the potential associations of phytobeneficial traits in PGPR bacteria, and this can now be achieved based on genome comparisons and phylogenetic analyses[17,18].

Hence, our objective was to assess the distribution of 23 genes contributing to eight key plant-beneficial functions using genomic and phylogenetic analyses, as well as ancestral state character reconstruction to infer possible gene transfers. These plant-beneficial function contributing genes (hereafter referred to as PBFC genes) were investigated using the genomes of 25 emblematic proteobacterial PGPR (i.e. bacteria colonizing root surface and/or tissues and displaying plant growth-promotion effects). These genomes were also compared with those of 279 other Alpha-, Beta- and Gammaproteobacteria representing various taxonomic groups and ecological status, such as (i) endophytes/symbionts (i.e. asymptomatic, endophytic bacteria possibly in symbiotic interaction with the plant but for which plant-beneficial effects are not documented, as well as root-nodulating, diazotrophic bacteria), (ii) saprophytes (i.e. bacteria from various environments including soil; some of them possibly colonizing roots but without established plant-beneficial effects), (iii) plant pathogens and (iv) animal pathogens.

The 23 genes selected included (i) the nitrogenase-encoding genes *nifHDK* responsible for nitrogen fixation in proteobacterial PGPR from *Azospirillum*[11], *Burkholderia*[19] and other genera, (ii) the pyrroloquinoline quinone-encoding genes *pqqBCDEFG* contributing to mineral phosphate solubilization in the PGPR *Pseudomonas fluorescens* F113[20], *Erwinia herbicola*[21] and *Enterobacter intermedium*[22], (iii) the indole-3-pyruvate decarboxylase/phenylpyruvate decarboxylase gene *ipdC*/*ppdC* of the indole-3-pyruvate pathway for synthesis of the auxinic phytohormone indole acetic acid (IAA) in *Azospirillum brasilense* Sp245[15], *Enterobacter cloacae* UW5[23] and other *Enterobacteriaceae* PGPR[24], (iv) the copper nitrite reductase gene *nirK* leading to formation of the NO root-branching signal in *Azospirillum brasilense* Sp245[25], (v) the 1-aminocyclopropane-1-carboxylate (ACC) deaminase gene *acdS* in *Pseudomonas putida* GR12-2[26] and various other *Pseudomonas* PGPR[6], which enables degradation of the plant's ethylene precursor, (vi) the acetoine genes *budAB* and 2,3-butanediol gene *budC* (induced systemic resistance) in the PGPR *Enterobacter* sp. 638[27], and (vii) genes *hcnABC* (hydrogen cyanide) and *phlACBD* (2,4-diacetylphloroglucinol) for synthesis of antimicrobial compounds in *P. fluorescens* F113, *P. protegens* CHA0 and many other PGPR pseudomonads[12].

Results

Contrasted co-occurrence patterns of PBFC genes in proteobacterial PGPR. In the 25 sequenced PGPR strains, which belonged to the genera *Azospirillum*, *Rhizobium*/*Agrobacterium* (Alphaproteobacteria), *Azoarcus*, *Burkholderia*, (Betaproteobacteria), and *Enterobacter*, *Klebsiella*, *Pantoea, Pseudomonas, Serratia*, (Gammaproteobacteria), the PBFC genes were found in 2 (for gene *ppdC*) to 20 (*pqqBCDE*) of the genomes (Table 1). The PGPR strains harbored from 1 (i.e. *acdS* in *Burkholderia* '*cepacia*' 383 and *B. phytofirmans* PSJN) to 14 of the 23 PBFC genes studied (in *P. protegens* Pf-5, *P. brassicacearum* NFM421 and *P. fluorescens* F113), which gave 7.5 ± 3.1 PBFC genes per strain (Supplementary Fig. S1a). The exact test of Fisher ($P < 0.05$) evidenced that *phlACBD* and *hcnABC* significantly occurred together in certain PGPR strains (Fig. 1) i.e. pseudomonads. Three other separate groups of co-occurring genes were identified, i.e. *budAB* and *ipdC*, the operon *nifHDK* and the clustered genes *pqqBCDE*. No other significant co-occurrence of PBFC genes was found.

Similar or lower prevalence of PBFC genes in Proteobacteria of other ecological types. The genomes of 279 other sequenced Proteobacteria corresponding to saprophytes or endophytes/symbionts without established PGPR status, as well as pathogens of plants or animals, were studied as well. For the 56 endophytes/symbionts, PBFC genes were found in 0 (for *phlACBD*) to 36 (*pqqBCDE*) of the genomes (Table 1). Whereas two bacteria did not display any of the 23 PBFC genes, they were extensively found in others, with eight strains exhibiting as many as 10 PBFC genes each. Overall, the endophytes/symbionts harbored 6.1 ± 2.6 PBFC genes per strain, but the difference with PGPR was not significant ($P = 0.06$). Exact-Fisher pairwise tests of the co-occurrence of PBFC genes ($P < 0.05$) revealed four groups, i.e. *hcnABC* and *pqqBCDE* linked by *pqqFG* genes, as well as *nifHDK*/*acdS* and *budAB*/*ipdC* further apart (Fig. 2a).

Within the 29 saprophytes, PBFC genes were found in 0 (for *phlACBD* and *ppdC*) to 26 (*pqqBCDE*) of the genomes (Table 1). Although three bacterial strains showed none of the studied genes, one strain (*Pantoea* sp. At-9b) exhibited as many as 12 genes. Globally, saprophytic strains contained 5.5 ± 1.8 genes per genomes. This is significantly lower than in PGPR ($P < 0.05$) but not different from endophytes/symbionts ($P = 0.44$). Co-occurrence analysis of PBFC genes in saprophytic bacteria evidenced five separate groups, i.e. *hcnABC*, *pqqBCDE*, *pqqFG*, *nifHDK* and *budABC*/*ipdC* (Fig. 2b).

In the 59 phytopathogenic bacteria, PBFC genes were found in 0 (for the 8 genes *ppdC*, *phlACBD* and *hcnABC*) to 35 (*pqqCDE*) of the genomes (Table 1). Whereas seven phytopathogens (*Xylella* sp. and *Xanthomonas albilineans*) did not contain any of the 23 PBFC genes, as many as 8 PBFC genes occurred in *Erwinia* and *Pantoea* species. This gave overall 4.3 ± 2.2 PBFC genes per strain, which was lower than in PGPR and endophytes/symbionts ($P < 0.05$) but not significantly lower than in saprophytes ($P = 0.06$). Exact-Fisher pairwise tests ($P < 0.05$) of the co-occurrence of PBFC genes revealed two independent groups, i.e. *pqqBCDEFG* linked to *acdS* via *pqqG*, and *budABC*/*ipdC* with *nifHDK* (Fig. 2c).

Most PBFC genes were not prevalent in the 135 animal pathogens. Except *nirK* present in 109 of them, the other PBFC genes were not often found (ranging from 4 genomes for *pqqG*, *phlD* and *hcnABC* to 44 genomes for *acdS*) or not found at all (*nifHDK*, *ppdC*, *phlACB*). The number of PBFC genes varied from 0 (in 9 animal pathogens) to 9 (in 7 other animal pathogens), i.e. 1.8 ± 1.2 PBFC genes per strain overall, which was lower than for all other ecological types (all $P < 0.05$). Exact-Fisher pairwise tests ($P < 0.05$) evidenced a single group comprised of three subgroups extensively linked with one another, i.e. *budABC*/*ipdC*, *pqqBCDEF* and *hcnABC*/*pqqG*/*phlD* (Fig. 2d).

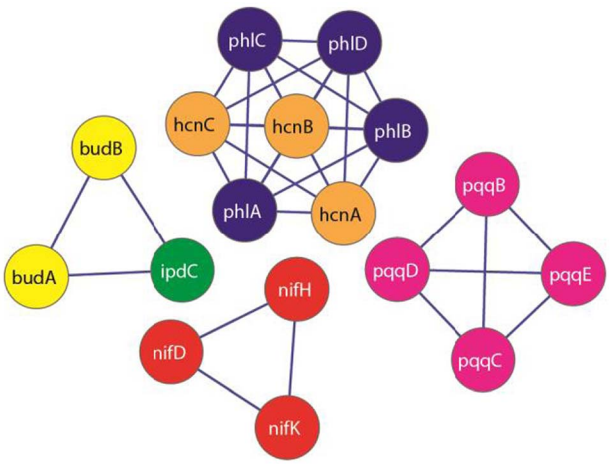

Figure 1 | Co-occurrence network of the PBFC genes for the 25 PGPR genomes. The genes are depicted with a colored circle according to their encoded function. Each co-occurrence is represented by an edge linking the corresponding genes and materialized by a line (based on Fisher exact test; $P < 0.05$). Several PBFC genes found in PGPR (i.e. *pqqF*, *pqqG*, *budC*, *nirK*, *ppdC* and *acdS*) did not display significant co-occurrence with any other(s).

Distribution of PBFC genes across all 304 proteobacterial genomes reveals taxonomic specificities. Whereas *phlACB* were only retrieved in PGPR (in 3 of 25 genomes), the other PBFC genes were recovered in bacteria of different ecological types. Many occurred in PGPR as well as in endophytes/symbionts, saprophytes and phytopathogens, especially *pqqCDE* (36 of 56, 26 of 29 and 35 of 59 genomes, respectively), and with a lower prevalence *ipdC* (2 of 56, 2 of 29 and 10 of 59 genomes, respectively) and *nifHDK* (23 of 56, 3 of the 29, and 3 of 59 genomes, respectively). In contrast, the *hcnABC* genes were retrieved in PGPR (3 of 25 genomes), saprophytes (2 of 29 genomes), endophytes/symbionts (9 of 59 genomes) and animal pathogens (4 of 135 genomes), but were absent in plant pathogens.

The distribution of certain PBFC genes according to bacterial ecological type could, at least in part, reflect taxonomic properties. This is indicated by the occurrence of PBFC genes in taxa restricted to a given ecological type (Fig. 3). In particular, *ppdC* was only retrieved in certain *Azospirillum* PGPR and *Bradyrhizobium* in the endophyte/symbiont category. For many PBFC genes, however, their occurrence within a taxon was related to species/strain ecology. This was the case for *phlACBD* (*Pseudomonas* PGPR), *hcnABC* (all *Pseudomonas* types except phytopathogens), and *nifHDK* (mainly in PGPR and endophytes/symbionts from various proteobacterial taxa). The relation to ecology, if any, was not as strong for *ipdC* and *budAB* (Enterobacteriaceae), *acdS* (all Burkholderiaceae considered and various Alphaproteobacteria and Gammaproteobacteria), *nirK* and *pqq* genes (various Proteobacteria corresponding to several ecological types).

The comparison of the 304 genomes showed that, unexpectedly, PBFC genes previously described as clustered (even forming operons in many cases) were not necessarily found together in a same genome (Fig. 4). For instance, *pqqFG* were close to *pqqBCDE* in *Pseudomonas* (and a few other genera), whereas *pqqBCDE* occurred without *pqqG* (encoding a family-S9 peptidase) and especially *pqqF* (encoding a family-M16 peptidase) in most other Proteobacteria. Similar observations were made for *phlD* and *phlACB*, as well as *budAB* and *budC*. Yet, the groups revealed by exact-Fisher pairwise tests ($P < 0.01$) corresponded mainly to genes involved in a same function (Fig. 3). This analysis showed that *hcnABC* and *phlD* linked the other *phl* genes with the six *pqq* genes, themselves linked to *budABC*/*ipdC* via *nirK* and to *nifHDK*. *nifHDK* were also linked, separately, to *ppdC* and to *acdS*.

Distribution of PBFC genes is partly related to proteobacterial phylogeny. We assessed whether the distribution of PBFC genes exhibited significant phylogenetic signal, meaning that closely-related species have similar gene content. Fritz and Purvis D index analysis (Table 2) showed that distribution of the PBFC genes was significantly influenced by evolutionary relationships between proteobacterial species, as indicated by D scores significantly less than 1. The genes *phlACBD*, *pqqFG*, *budABC*, *ipdC*, *ppdC* and *hcnABC* showed a strong phylogenetic signal, while *acdS*, *nifHDK* and *pqqBCDE* showed weaker signals.

Horizontal gene transfers had significant effects on PBFC gene distribution in Proteobacteria. When the impact of genome plasticity was assessed, by computing events of acquisitions and losses across proteobacterial species, no loss was detected for *pqqF*, *phlACBD*, *ppdC* and *nirK* (Table 3). On the contrary, a few losses were inferred for the other genes, ranging from 1 loss for *pqqG*, *budABC*, *ipdC* and *hcnABC* to 6 losses for *pqqB*. In comparison,

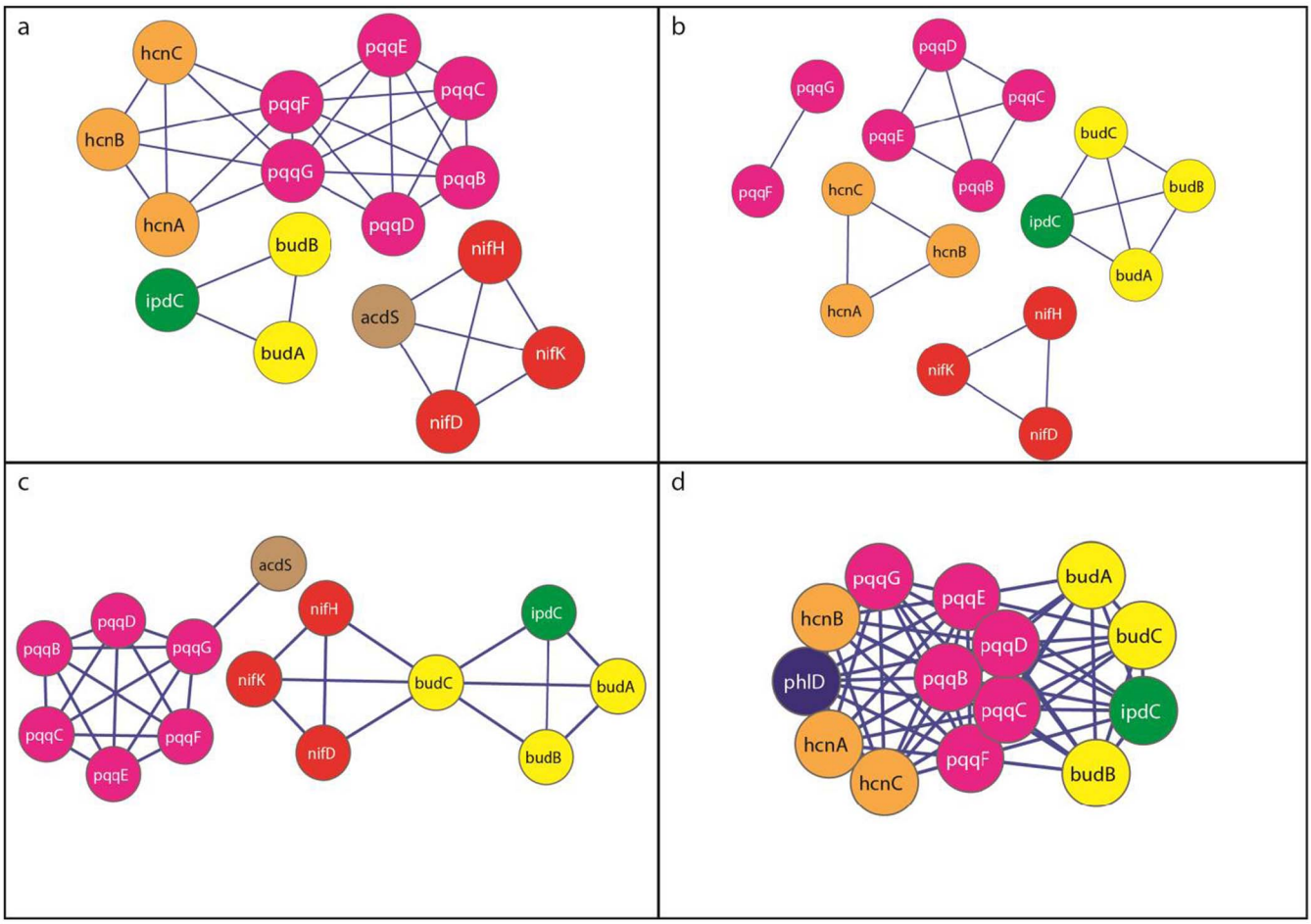

Figure 2 | Co-occurrence network of PBFC genes according to primary ecological classification of bacteria. The genes are depicted with a colored circle according to their encoded function. Each co-occurrence is represented by an edge linking the corresponding genes and materialized by a line. Computations were made for (a) endophytes/symbionts, (b) saprophytes, (c) phytoparasites, (d) animal pathogens.

the number of acquisitions was of a larger scale, from 1 for *ipdC* and *budAB* to 21 for *acdS* (Table 3).

All 23 genes appeared at least once in a distant ancestor of the species studied (Fig. 5). *ipdC*, *ppdC* and *phlACBD* are clade specific; *ipdC* appeared in the last common ancestor (LCA) of *Pantoea* and *Erwinia* genera, *ppdC* in the LCA of *Azospirillum brasilense* and the LCA of *Bradyrhizobium* strains ORS78 and BTAi1, and *phlACBD* in the LCA of *Pseudomonas fluorescens* F113 and *Pseudomonas brassicacearum* NFM421 and the LCA of *Pseudomonas protegens* Pf-5. *budABC* appeared in the LCA of Enterobacteriaceae; *budAB* are clade specific but *budC* was acquired at least 7 times in other clades. The *pqqBCDEFG* genes appeared in the LCA of the *Pseudomonas*; *pqqG*, *pqqF* and *pqqBCDE* were acquired respectively 4, 5 and 15 times by other taxa. At the extreme, *nifHDK* underwent at least 18 acquisitions and *acd*S (which appeared in the Burkholderiaceae LCA) 21 acquisitions, in both cases across the three phyla considered.

Discussion

In this study, plant-beneficial properties of PGPR were for the first time assessed on a broad scale, by considering (i) a large range of PBFC genes corresponding to various types of plant-beneficial properties, (ii) PGPR strains of contrasted taxonomic status (from the Alpha- Beta- and Gammaproteobacteria), and (iii) a selection of non-PGPR Proteobacteria with primarily other biotic relations with plants (i.e. endophytes/symbionts and phytopathogens) or other types of ecology (i.e. saprophytes and animal pathogens).

It could have been thought that the PGPR status entailed the presence of a core collection of PBFC genes shared by all PGPR strains, but the current results based on 25 emblematic PGPR strains indicate that none of the 23 key PBFC genes of the study were common to all strains, even though as many as 20 PGPR genomes displayed *pqqBCDE*. PQQ is a co-factor potentially implicated in several cellular processes (and incidentally contributing also to phosphate solubilization), which may explain its wide occurrence in PGPR[13,28]. In comparison with Proteobacteria of other lifestyle, PBFC genes restricted to PGPR were not found, except for *phlACB* but these genes were present in only 3 *Pseudomonas* PGPR. However, the number of PBFC genes increased along the continuum animal pathogens (only 1.8 PBFC genes/strain), phytopathogens, saprophytes, endophytes/symbionts, PGPR (as many as 7.5 PBFC genes/strain). The same findings were made when assessing the number of functions expected from these PBFC genes, except that the difference between animal and plant pathogens was not significant (Supplementary Fig. S1b).

Our gene distribution data suggest that PBFC genes might by selected in plant-associated habitats and counter-selected elsewhere, as exemplified by the very low number of these genes in animal pathogens (where only *nirK* was prevalent). This is in accordance with the expectations that most of the corresponding functions would not be relevant for animal physiology and plant is not the primary habitat of these bacteria. For instance, nitrogen fixation is counter-selected in pathogenic bacteria[29,30]. In addition, results suggest that amongst all the plant-associated bacteria, specific lifestyle is

Figure 3 | Phylogenetic distribution of genes along Proteobacteria phylogeny. Internal circles: presence of a gene is indicated by a grey square and absence by a white square. Taxonomically coherent groups with the same gene content were collapsed for sake of clarity. Biovars are indicated for *Rhizobium leguminosarum* and pathovars for *Pseudomonas syringae*.

also a major factor explaining distribution of PBFC genes, with higher prevalence in plant-beneficial strains. This possibility stems in particular from the comparison of (i) PGPR and endophytes/symbionts versus (ii) phytopathogens, despite the presence of PBFC homologs *budABC*, *ipdC* and/or *acdS* (not necessarily together; Fig. 2c) in many phytopathogens (Table 1). Indeed, many of the plant-beneficial traits found in PGPR could be used by endophytic Proteobacteria documented (or presumably) in a mutualistic symbiosis with the plant host. This would be a generalization of previous observations made with *nifHDK*[31] and to a lesser extent *acdS*[32].

Most PBFC genes were identified in bacteria from different ecological types (Table 1), which is an indication that (i) strain information was not always sufficient to determine lifestyle precisely and/or (ii) boundaries between different lifestyles may not be very stringent in Proteobacteria. The first possibility is clear in the case of saprophytes, as this category contains a number of strains originating from bulk or rhizosphere soil but for which the PGPR potential has not been experimentally tested, raising the possibility that some of them could indeed be PGPR. Similarly, certain PGPR can also be endophytic, e.g. *Azospirillum* sp. B510 and *Azoarcus* sp. BH72[38], but some of the endophytes studied here have not been assessed for their effects on plants and so could not be listed among the PGPR. The second possibility is illustrated with many animal pathogens belonging to *Pseudomonas*[33] or the Enterobacteriaceae[34] that can colonize plants

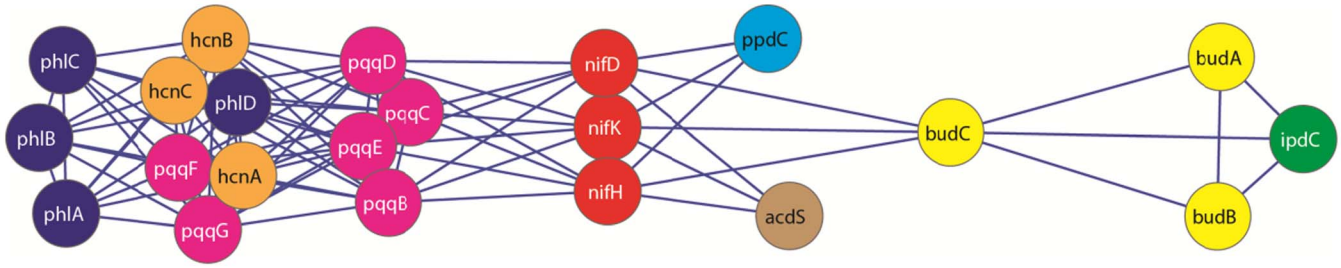

Figure 4 | Co-occurrence network of the PBFC genes for the 304 genomes. The genes are depicted with a colored circle according to their encoded function. Each co-occurrence is represented by an edge linking the corresponding genes and materialized by a line. *nirK* does not appear in the figure because this gene did not shown any significant co-occurrence with other PBFC gene(s).

asymptomatically, probably because these alternative hosts promote bacterial survival before recolonizing the next animal primary host[35]. This could explain why certain animal-associated strains displayed PBFC genes. Furthermore, opportunistic human pathogens such as *Pseudomonas aeruginosa* PAO1 and PA14 can also infect roots and lead to plant death[36].

Mutational inactivation of a particular PBFC gene may reduce (without necessarily abolishing) plant-beneficial effects in PGPR strains[1,37], and genetic acquisition of an additional PBFC gene has the potential to enhance PGPR performance[8,39]. This indicates that possessing multiple PBFC genes should confer a better efficiency at enhancing plant growth. In this context, the analysis of co-occurrence patterns (exact Fisher tests) can be useful to identify selection of multiple PBFC genes and their potential synergistic effects. However, gene co-occurrence may also take place because species that share a recent evolutionary history also share similar gene contents, a phenomenon known as phylogenetic signal. Indeed, the Fritz and Purvis index clearly pointed to gene associations related to phylogenetic signal, i.e. PBFC genes were more likely to be conserved in closely-related species. This also raises the possibility that the potential to become a PGPR may rely (at least in part) on ancestral features in the corresponding bacterial taxa, which is in phase with previous findings on particular PGPR populations[40] and more generally on function distributions in Gammaproteobacteria[41].

The distribution pattern of PBFC genes amongst Proteobacteria of various lifestyles and the relation to bacterial taxonomy prompted us to assess in more details the evolutionary history of these genes. Ancestral character reconstructions showed few losses of PBFC genes, even in animal-associated bacteria, but many more gene acquisitions. Indeed, the role of horizontal gene transfer has been substantiated with various types of PGPR[42,43] and suggests that cooperation interactions between Proteobacteria and plant roots might have established separately in various taxa, yielding PGPR strains whose effect(s) on the plant may rely on different and taxa-specific combinations of modes of action. Further genome sequencing efforts targeting close relatives of these PGPR would be needed to confirm this possibility. Despite conservation of PBFC genes across different ecological lifestyles, a differential use/regulation of these genes depending on environmental and host conditions is likely[44], as can take place during exaptation[45]. Indeed, expression patterns of PBFC

Table 2 | Phylogenetic patterns of gene distribution in selected Proteobacteria. Values were calculated for the 1000 partitions of the species phylogenetic tree

Gene[a]	*D*[b]	*P(D > 0)*[c]	*P(D < 1)*[c]	*Phylogenetic signal strength*
pqqB	0.05 (0.01/0.07)	0.16/0.41	0	Strong
pqqC	0.04 (0.00/0.07)	0.16/0.45	0	Strong
pqqD	0.04 (0.00/0.07)	0.16/0.45	0	Strong
pqqE	0.04 (0.01/0.07)	0.17/0.45	0	Strong
pqqF	−0.17 (−0.20/−0.15)	0.95/0.99	0	Very strong
pqqG	−0.18 (−0.21/−0.17)	0.95/0.98	0	Very strong
phlA	−0.40 (−0.76/−0.06)	0.48/0.88	0.00/0.04	Very strong
phlB	−0.40 (−0.75/−0.05)	0.49/0.89	0.00/0.04	Very strong
phlC	−0.40 (−0.75/−0.06)	0.48/0.88	0.00/0.03	Very strong
phlD	−0.24 (−0.36/−0.12)	0.63/0.86	0	Very strong
hcnA	−0.18 (−0.30/−0.06)	0.62/0.95	0	Strong
hcnB	−0.18 (−0.30/−0.05)	0.61/0.95	0	Strong
hcnC	−0.18 (−0.29/−0.06)	0.61/0.94	0	Strong
budA	−0.34 (−0.37/−0.30)	0.98/0.99	0	Very strong
budB	−0.34 (−0.37/−0.30)	0.98/0.99	0	Very strong
budC	−0.02 (−0.08/−0.03)	0.40/0.70	0	Strong
nirK	−0.05 (−0.09/−0.02)	0.60/0.89	0	Very strong
ipdC	−0.33 (−0.41/0.28)	0.96/0.99	0	Very strong
ppdC	−0.34 (−0.27/−0.47)	0.65/0.80	0	Very strong
acdS	0.10 (0.07/0.13)	0.05/0.19	0	Moderate
nifD	0.28 (0.23/0.33)	0.01/0.05	0	Weak
nifH	0.28 (0.24/0.33)	0.00/0.04	0	Weak
nifK	0.28 (0.23/0.33)	0.00/0.04	0	Weak

[a]The genes studied are involved in phosphate solubilization (pyrroloquinoline quinone; *pqqBCDEFG*), 2,4-diacetylphloroglucinol synthesis (*phlACBD*), hydrogen cyanide synthesis (*hcnABC*), induced systemic resistance (acetoine and 2,3-butanediol; *budAB* and *budC*, respectively), NO synthesis (copper nitrite reductase; *nirK*), IAA synthesis (indole-3-pyruvate decarboxylase/phenylpyruvate decarboxylase; *ipdC/ppdC*), plant ethylene regulation (ACC deamination; *acdS*), and nitrogen fixation (nitrogenase; *nifHDK*).
[b]Median value (with the minimum and maximum values in parenthesis).
[c]Minimum and maximum values when different.

Table 3 | Number of acquisitions and losses for each PBFC gene[a] according to proteobacterial species tree (ancestral character reconstruction)

Acquisitions and losses	*pqqB*	*pqqC*	*pqqD*	*pqqE*	*pqqF*	*pqqG*	*phlA*	*phlB*	*phlC*	*phlD*	*hcnA*	*hcnB*	*hcnC*	*budA*	*budB*	*budC*	*nirK*	*ipdC*	*ppdC*	*acdS*	*nifD*	*nifH*	*nifK*
Acquisitions	16	16	16	16	4	3	2	2	2	2	4	4	4	1	1	8	11	1	2	21	18	18	18
Losses	6	5	5	5	0	1	0	0	0	0	1	1	1	1	1	1	0	1	0	4	2	2	2
Total occurrences	129	129	129	129	58	50	3	3	3	7	18	18	18	29	29	42	131	25	4	125	38	38	38

[a]The genes studied are involved in phosphate solubilization (pyrroloquinoline quinone; *pqqBCDEFG*), 2,4-diacetylphloroglucinol synthesis (*phlACBD*), hydrogen cyanide synthesis (*hcnABC*), induced systemic resistance (acetoine and 2,3-butanediol; *budAB* and *budC*, respectively), NO synthesis (copper nitrite reductase; *nirK*), IAA synthesis (indole-3-pyruvate decarboxylase/phenylpyruvate decarboxylase; *ipdC/ppdC*), plant ethylene regulation (ACC deamination; *acdS*), and nitrogen fixation (nitrogenase; *nifHDK*).

genes according to taxonomic and/or lifestyle properties is an important ecological issue, which will deserve further research attention. Bacterial adaptation to new niches is mainly dependent on genetic novelty[46,47], which may entail gene acquisitions[46] or differential regulation[48]. Many examples of traits conferring environmental adaptation that were further co-opted as virulence factor are documented in human pathogens[49]. Similar processes are likely to have taken place in PGPR as well[50].

In conclusion, the comparison of taxonomically-contrasted proteobacterial PGPR with a wide range of related, non-PGPR bacteria suggested that the emergence of the PGPR status could have paralleled accumulation of PBFC genes in root-adapted bacteria. It is likely that this process took place separately in taxonomically-contrasted Proteobacteria and involved ancient gene acquisitions, which explains why subsequent diversification produced taxonomic subgroups of PGPR strains differing from one another in the range of PBFC genes accumulated.

Methods

Selection of genomes. The genomes used were selected among those available in October 2012. They corresponded to 25 PGPR, 56 endophytes/symbionts (35 endophytes and 21 root-nodulating bacteria), 29 saprophytes (3 from water environments, 6 from bulk soil, 16 from the rhizosphere, and 4 from healthy animal samples), 59 plant pathogens and 135 animal pathogens (124 of them infecting humans). Since distribution can be influenced by phylogenetic relatedness, also called phylogenetic signal[18], genomes were chosen so as to balance the prevalence of the various Alpha-, Beta- and Gammaproteobacterial groups for which PGPR genomes were available, following two principles. First, the primary lifestyle of the selected bacteria had to be documented sufficiently clearly and their genomes were fully sequenced (except in a few cases for orders of particular interest). Second, bacterial orders in which PGPR representatives were available were assessed for genome availability of bacteria corresponding to other lifestyles (especially within the same or closely-related families/genera), and if unsuccessful the phylogenetically-closest order was then targeted.

Homologs retrieval. Homologs of genes contributing to a phytobeneficial function in PGPR were retrieved using a BLAST-based method. A protein to protein search was done using Blastp[51] with a subset of genes documented to contribute to a given phytobeneficial function (Supplementary Table S1). As annotations in public databases may contain errors or sometimes fail to accurately predict gene identity, we then did a tblastn[51] search on genomic sequences to overcome these limitations. An E-value threshold of 1e-15 was set to filter blast searches.

Protein family assignment. Assignment of homologous proteins to families having the same putative function was done using a combination of significant sequence identity (see above) and protein domain assignment. Protein domain assignment was done using rps-blast[52] and the Conserved Domain Database (CDD)[53]. We separated the NCBI-curated domains (which are considered more accurate) and external sources domains in two distinct databases. The NCBI-curated database was preferentially used for protein domain assignments while external source database was used when the NCBI-curated one could not retrieve results. Proteins were considered of the same family if they (i) had at least 30% of identity on at least 70% of their respective protein sequence length and (ii) shared the same domains with a reference phytobeneficial protein. Phylogenetic profiles (corresponding to a binary vector with gene's presence and absence respectively indicated as 1 and 0 for each genome) were used to represent the presence/absence of a particular gene in the different organisms for analysis of phylogenetic signal and ancestral state reconstruction.

Gene distribution. For statistical analysis of the number of PBFC genes and number of corresponding functions per genome, according to primary bacterial ecology, the Wilcoxon test was used with the R command *wilcox.test* ($P < 0.05$).

Proteobacterial phylogenetic tree. The proteobacterial phylogenetic tree was based on 31 housekeeping markers identified, aligned and trimmed with Amphora2, as done previously[54]. Trees were inferred by ExaML[55] with the concatenated alignment, 1000 replicates and the PSR model of rate heterogeneity.

Computation of phylogenetic signal. The phylogenetic signal for each gene was calculated using Fritz and Purvis's D index[56] implemented in the R package "caper". Computation of random and Brownian motion of evolution probabilities was based on 10,000 permutations. Briefly, a given trait (a gene in our case) displays a highly clustered distribution if $D < 0$, is as clustered as if it evolved under Brownian motion

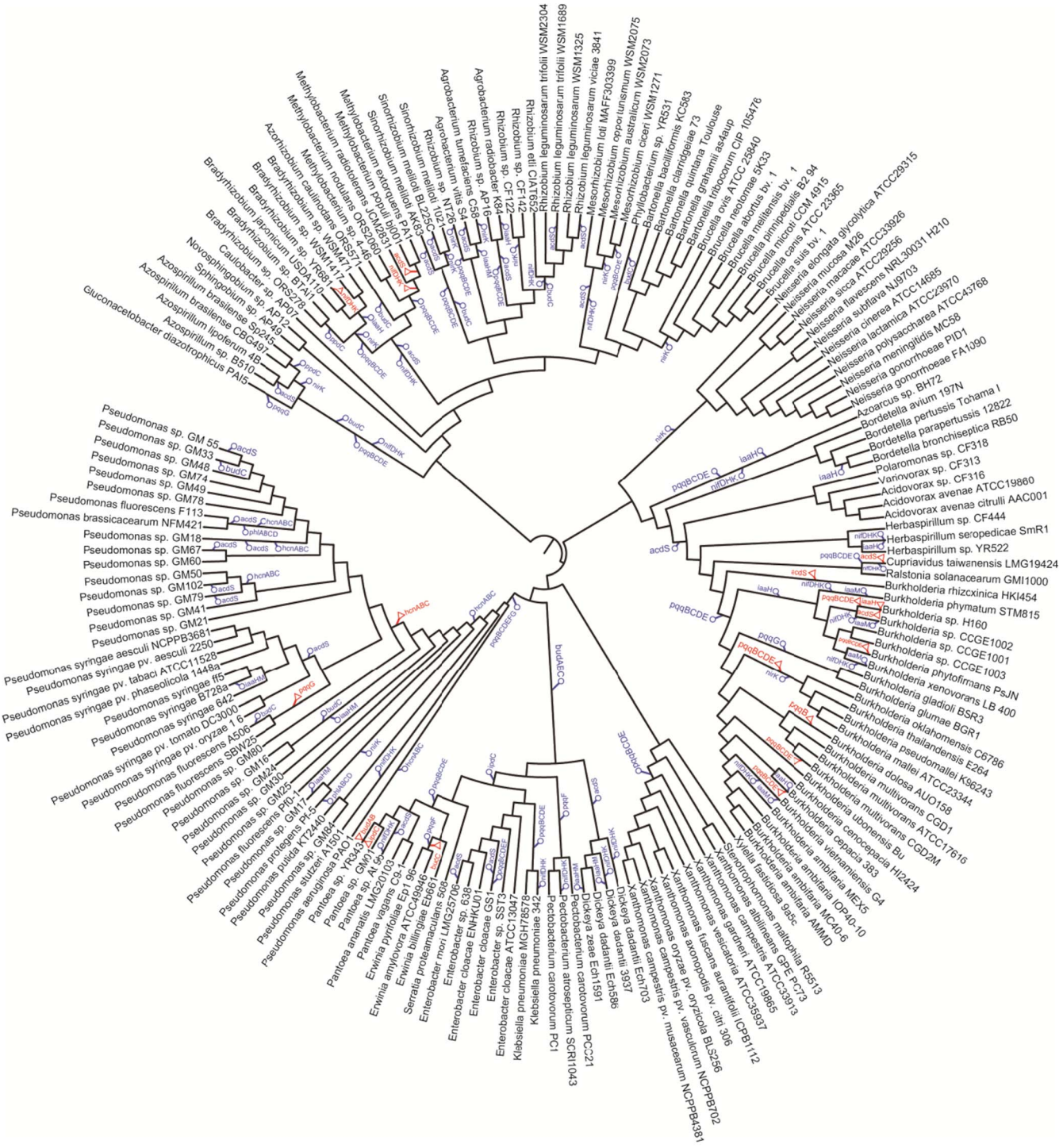

Figure 5 | Reconstruction of acquisitions and losses of PBFC genes in relation to evolutionary history of sequenced bacteria. When different strains of the same species had the same PBFC gene profile, only one representative strain was kept in the Maximum-Likelihood tree to avoid redundant information. Acquisitions are indicated by a blue arrow with a circle and losses by a red arrow with a triangle.

if D = 0, displays random distribution if D = 1, and is overdispersed if D > 1. Comparison of D scores was used to arbitrarily infer the strength of the phylogenetic signal for each gene.

Ancestral state character reconstruction. The GLOOME algorithm was used to infer the presence or absence of each gene on each node of a phylogenetic tree based on their distributions in terminal taxa. The phylogenetic tree used was computed as previously but was based on a filtered alignment. When many bacteria of the same species had the same content in genes of interest, only the reference species indicated in the NCBI database was conserved. This simplified the reconstruction model by removing redundant information. Reconstructions were made with the Maximum Parsimony method[57], which allows to reconstruct ancestral states by minimizing character change events along a phylogenetic tree.

1. Vacheron, J. *et al.* Plant growth-promoting rhizobacteria and root system functioning. *Front Plant Sci* **4**, 356 (2013).
2. Lugtenberg, B. & Kamilova, F. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* **63**, 541–556 (2009).
3. Drogue, B., Doré, H., Borland, S., Wisniewski-Dyé, F. & Prigent-Combaret, C. Which specificity in cooperation between phytostimulating rhizobacteria and plants? *Res Microbiol* **163**, 500–510 (2012).
4. Somers, E., Vanderleyden, J. & Srinivasan, M. Rhizosphere bacterial signalling: a love parade beneath our feet. *Crit Rev Microbiol* **30**, 205–240 (2004).

5. Glick, B. R. Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *FEMS Microbiol Lett* **251**, 1–7 (2005).
6. Blaha, D., Prigent-Combaret, C., Mirza, M. S. & Moënne-Loccoz, Y. Phylogeny of the 1-aminocyclopropane-1-carboxylic acid deaminase-encoding gene *acdS* in phytobeneficial and pathogenic Proteobacteria and relation with strain biogeography. *FEMS Microbiol Ecol* **56**, 455–470 (2006).
7. Zamioudis, C. & Pieterse, C. M. J. Modulation of host immunity by beneficial microbes. *Mol Plant-Microbe Interact* **25**, 139–150 (2012).
8. Haas, D. & Défago, G. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat Rev Microbiol* **3**, 307–319 (2005).
9. Couillerot, O., Prigent-Combaret, C., Caballero-Mellado, J. & Moënne-Loccoz, Y. *Pseudomonas fluorescens* and closely-related fluorescent pseudomonads as biocontrol agents of soil-borne phytopathogens. *Lett Appl Microbiol* **48**, 505–512 (2009).
10. Walker, V. *et al.* Unexpected phytostimulatory behavior for *Escherichia coli* and *Agrobacterium tumefaciens* model strains. *Mol Plant-Microbe Interact* **26**, 495–502 (2013).
11. Bashan, Y. & de-Bashan, L. E. How the plant growth-promoting bacterium *Azospirillum* promotes plant growth - a critical assessment. *Adv Agron* **108**, 77–136 (2010).
12. Haas, D. & Keel, C. Regulation of antibiotic production in root-colonizing P*seudomonas* spp. and relevance for biological control of plant disease. *Annu Rev Phytopathol* **41**, 117–153 (2003).
13. Misra, H. S., Rajpurohit, Y. S. & Khairnar, N. P. Pyrroloquinoline-quinone and its versatile roles in biological processes. *J Biosci* **37**, 313–325 (2012).
14. Ellis, R. J., Timms-Wilson, T. M. & Bailey, M. J. Identification of conserved traits in fluorescent pseudomonads with antifungal activity. *Env Microbiol* **2**, 274–284 (2000).
15. Spaepen, S. *et al.* Characterization of phenylpyruvate decarboxylase, involved in auxin production of *Azospirillum brasilense*. *J Bacteriol* **189**, 7626–7633 (2007).
16. Almario, J., Gobbin, D., Défago, G., Moënne-Loccoz, Y. & Rezzonico, F. Prevalence of type III secretion system in effective biocontrol pseudomonads. *Res Microbiol* **165**, 300–304 (2014).
17. Kim, P.-J. & Price, N. D. Genetic co-occurrence network across sequenced microbes. *PLoS Comp Biol* **7**, e1002340 (2011).
18. Martiny, A. C., Treseder, K. & Pusch, G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**, 830–838 (2013).
19. Suárez-Moreno, Z. R. *et al.* Common Features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microb Ecol* **63**, 249–266 (2012).
20. Miller, S. H. *et al.* Biochemical and genomic comparison of inorganic phosphate solubilization in *Pseudomonas* species. *Env Microbiol Rep* **2**, 403–411 (2010).
21. Liu, S. T. *et al.* Cloning of an *Erwinia herbicola* gene necessary for gluconic acid production and enhanced mineral phosphate solubilization in *Escherichia coli* HB101: nucleotide sequence and probable involvement in biosynthesis of the coenzyme pyrroloquinoline quinone. *J Bacteriol* **174**, 5814–5819 (1992).
22. Kim, C. H. *et al.* Cloning and expression of pyrroloquinoline quinone (PQQ) genes from a phosphate-solubilizing bacterium *Enterobacter intermedium*. *Curr Microbiol* **47**, 457–461 (2003).
23. Ryu, R. J. & Patten, C. L. Aromatic amino acid-dependent expression of indole-3-pyruvate decarboxylase is regulated by TyrR in *Enterobacter cloacae* UW5. *J Bacteriol* **190**, 7200–7208 (2008).
24. Zimmer, W., Hundeshagen, B. & Niederau, E. Demonstration of the indolepyruvate decarboxylase gene homologue in different auxin-producing species of the Enterobacteriaceae. *Can J Microbiol* **40**, 1072–1076 (1994).
25. Creus, C. M. *et al.* Nitric oxide is involved in the *Azospirillum brasilense*-induced lateral root formation in tomato. *Planta* **221**, 297–303 (2005).
26. Glick, B. R., Jacobson, C. B., Schwarze, M. M. K. & Pasternak, J. J. 1-Aminocyclopropane-1-carboxylic acid deaminase mutants of the plant growth promoting rhizobacterium *Pseudomonas putida* GR12-2 do not stimulate canola root elongation. *Can J Microbiol* **40**, 911–915 (1994).
27. Taghavi, S. *et al.* Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp. 638. *PLoS Genet* **6**, e1000943 (2010).
28. Choi, O. *et al.* Pyrroloquinoline quinone is a plant growth promotion factor produced by *Pseudomonas fluorescens* B16. *Plant Physiol* **146**, 657–668 (2008).
29. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of nitrogen fixation. *Mol Biol Evol* **21**, 541–554 (2004).
30. Menard, A. *et al.* Selection of nitrogen-fixing deficient *Burkholderia vietnamiensis* strains by cystic fibrosis patients: involvement of *nif* gene deletions and auxotrophic mutations. *Env Microbiol* **9**, 1176–1185 (2007).
31. Spaepen, S., Vanderleyden, J. & Okon, Y. Plant growth-promoting actions of rhizobacteria. *Adv Bot Res* **51**, 283–320 (2009).
32. Prigent-Combaret, C. *et al.* Physical organization and phylogenetic analysis of *acdR* as leucine-responsive regulator of the 1-aminocyclopropane-1-carboxylate deaminase gene *acdS* in phytobeneficial *Azospirillum lipoferum* 4B and other Proteobacteria. *FEMS Microbiol Ecol* **65**, 202–219 (2008).
33. Mark, G. L. *et al.* Transcriptome profiling of bacterial responses to root exudates identifies genes involved in microbe-plant interactions. *Proc Natl Acad Sci USA* **102**, 17454–17459 (2005).
34. Tyler, H. L. & Triplett, E. W. Plants as a habitat for beneficial and/or human pathogenic bacteria. *Annu Rev Phytopathol* **46**, 53–73 (2008).
35. Holden, N., Pritchard, L. & Toth, I. Colonization outwith the colon: plants as an alternative environmental reservoir for human pathogenic enterobacteria. *FEMS Microbiol Rev* **33**, 689–703 (2009).
36. Walker, T. S. *et al. Pseudomonas aeruginosa*-plant root interactions. pathogenicity, biofilm formation, and root exudation. *Plant Physiol.* **134**, 320–331 (2004).
37. Couillerot, O. *et al.* The role of the antimicrobial compound 2,4-diacetylphloroglucinol in the impact of biocontrol *Pseudomonas fluorescens* F113 on *Azospirillum brasilense* phytostimulators. *Microbiology* **157**, 1694–1705 (2011).
38. Krause, A. *et al.* Complete genome of the mutualistic, N_2-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nat Biotechnol* **24**, 1385–1391 (2006).
39. Baudoin, E. *et al.* Effects of *Azospirillum brasilense* with genetically modified auxin biosynthesis gene *ipdC* upon the diversity of the indigenous microbiota of the wheat rhizosphere. *Res Microbiol* **161**, 219–226 (2010).
40. Frapolli, M., Pothier, J. F., Défago, G. & Moënne-Loccoz, Y. Evolutionary history of synthesis pathway genes for phloroglucinol and cyanide antimicrobials in plant-associated fluorescent pseudomonads. *Mol Phylogenet Evol* **63**, 877–890 (2012).
41. Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**, e130 (2005).
42. Rezzonico, F., Défago, G. & Moënne-Loccoz, Y. Comparison of ATPase-encoding type III secretion system *hrcN* genes in biocontrol fluorescent pseudomonads and in phytopathogenic proteobacteria. *Appl Environ Microbiol* **70**, 5119–5131 (2004).
43. Wisniewski-Dyé, F. *et al. Azospirillum* genomes reveal transition of bacteria from aquatic to terrestrial environments. *PLoS Genet* **7**, e1002430 (2011).
44. Marchetti, M. *et al.* Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol* **8**, e1000280 (2010).
45. Gould, S. J. & Vrba, E. Exaptation; a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
46. Treangen, T. J. & Rocha, E. P. C. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**, e1001284 (2011).
47. Andres, J. *et al.* Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-26. *Genome Biol Evol* **5**, 934–953 (2013).
48. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012).
49. Adiba, S., Nizak, C., van Baalen, M., Denamur, E. & Depaulis, F. From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. *PLoS ONE* **5**, e11882 (2010).
50. Hardoim, P. R., van Overbeek, L. S. & van Elsas, J. D. Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol* **16**, 463–471 (2008).
51. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
52. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
53. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41**, D348–D352 (2012).
54. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).
55. Stamatakis, A. & Aberer, A. J. Novel parallelization schemes for large-scale likelihood-based phylogenetic inference. in *2013 IEEE 27th International Symposium on Parallel Distributed Processing (IPDPS)* 1195–1204 (2013). doi:10.1109/IPDPS.2013.70.
56. Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv Biol* **24**, 1042–1051 (2010).
57. Cohen, O. & Pupko, T. Inference of gain and loss events from phyletic patterns using stochastic mapping and maximum parsimony—a simulation study. *Genome Biol Evol* **3**, 1265–1275 (2011).

Acknowledgments

This work was supported in part by the Ministère Français de la Recherche. We are grateful to Audrey Dubost and Danis Abrouk (iBio platform, UMR CNRS 5557 Ecologie Microbienne) for helpful discussion.

Author contributions

C.P.C., D.M. and Y.M.L. conceived and designed the project. M.B. performed sequence queries and bioinformatics analyses. M.B., C.P.C., D.M. and Y.M.L. analyzed the findings and wrote the manuscript. All of the authors approved the final manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Bruto, M., Prigent-Combaret, C., Muller, D. & Moënne-Loccoz, Y. Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related Proteobacteria. *Sci. Rep.* **4**, 6261; DOI:10.1038/srep06261 (2014).