

RESEARCH ARTICLE

Open Access

# Evolution of REP diversity: a comparative study

Jaroslav Nunvar<sup>1,2\*</sup>, Irena Licha<sup>1</sup> and Bohdan Schneider<sup>2</sup>

## Abstract

**Background:** Repetitive extragenic palindromic elements (REPs) constitute a group of bacterial genomic repeats known for their high abundance and several roles in host cells' physiology. We analyzed the phylogenetic distribution of particular REP classes in genomic sequences of sixty-three bacterial strains belonging to the *Pseudomonas fluorescens* species complex and ten strains of *Stenotrophomonas* sp., in order to assess intraspecific REP diversity and to gain insight into long-term REP evolution.

**Results:** Based on proximity to RAYT (REP-associated tyrosine transposase) genes, twenty-two and thirteen unique REP classes were determined in fluorescent pseudomonads and stenotrophomonads, respectively. In stenotrophomonads, REP elements were typically found in tens or a few hundred copies per genome. REPs of fluorescent pseudomonads were generally more numerous, occurring in hundreds or even over a thousand perfect copies of particular REP class per genome. REP sequences showed highly heterogeneous distribution. The abundances of REP classes roughly followed host strains' phylogeny, differing markedly among individual clades. High abundances of particular REP classes appeared to depend on the presence of the cognate RAYT gene, and deviations from this state could be attributed to recent or ancient mutations of *rayt*-flanking REPs, or RAYT loss. RAYTs of both studied bacterial groups are monophyletic, and their cognate REPs show species-specific characteristics, suggesting shared evolutionary history of REPs, RAYTs and their hosts.

**Conclusions:** The results of our large-scale analysis show that REP elements constitute intriguingly dynamic components of genomes of fluorescent pseudomonads and stenotrophomonads, and indicate that REP diversification and proliferation are ongoing processes. High numbers of REPs have apparently been retained during the entire evolutionary time since the establishment of these two bacterial lineages, probably because of their beneficial effect on host long-term fitness. REP elements in these bacteria represent a suitable platform to study the interplay between repeated elements, their mobilizers and host bacterial cells.

**Keywords:** REP elements, *Stenotrophomonas maltophilia*, *Pseudomonas fluorescens*

## Background

Genomes of many higher eukaryotes are known to teem with repetitive DNA elements. By contrast, bacteria are notorious for their high coding density [1], which leaves significantly less space for expansion of repeats. Repetitive elements identified in bacteria can be generally divided into coding and noncoding ones. The former is typically represented by insertion sequences and transposons, parasitic DNA elements that catalyze their own movement and replication (with help of host cell's functions) [2]. Noncoding repeats (apart from repeated genes coding for

structural RNAs) comprise several distinct types, often connected to various cellular functions. For example, short, overrepresented DNA motifs mark DNA to be taken up by natural transformation in *Haemophilus* and related bacteria [3]. Similarly, Chi sequences, which serve as sites of recombination initiation, are overrepresented in host genomes [4]. Repeated elements are part of sophisticated CRISPR systems, which provide defense against invading mobile elements [5]. Finally, various types of MITEs (miniature inverted repeats transposable elements), which are predicted to be derived from autonomous transposable elements, are implicated in transcription regulation and other processes [6,7].

REP (repetitive extragenic palindrome) elements have now been known for over 30 years [8], originally from *Escherichia coli* and related enterobacteria [9]. They

\* Correspondence: nunvar@natur.cuni.cz

<sup>1</sup>Department of Genetics and Microbiology, Faculty of Science, Charles University, Vinicna 5, 128 44, Prague 2, Czech Republic

<sup>2</sup>Institute of Biotechnology of the ASCR, v. v. i., Videnska 1083, 142 20, Prague 4, Czech Republic

were later identified in other species, belonging predominantly to gammaproteobacteria – *Pseudomonas putida* [10], *Pseudomonas fluorescens* [11,12], *Stenotrophomonas maltophilia* [13], *Xanthomonas campestris* and others [14], each species possessing different types of REP sequences. REPs are typically highly numerous and occur almost exclusively in intergenic regions. The definition of REP elements was recently refined [14] to reflect their common features on sequence level: a 5'-terminal conserved tetranucleotide (GT<sup>A</sup>/<sub>C</sub>G) and downstream complementary (palindromic) region with variable base composition. REP elements are mostly arranged into repeats of higher order. REPINs (REP doublet forming hairpin) are composed of two closely spaced REPs in inverted orientation [15] and were found to represent the predominant REP form in *P. fluorescens* [11,15], *P. putida* [10] and *S. maltophilia* [13]. BIMEs (bacterial interspersed mosaic elements), abundant in *E. coli*, consist of tandemly repeated REPIN-like doublets [16]. Importantly, in *E. coli*, three significant proteins interact with REPs or BIMEs: integration host factor [17], DNA gyrase [18] and DNA polymerase I [19], indicating the role for these elements in major cellular processes. Furthermore, REPs were shown to modulate transcription and mRNA stability in both *E. coli* [20] and *S. maltophilia* [13]. REPs inhabit only the core parts of host genomes and are absent from laterally transferred regions [11-13].

A few years ago, we described a protein family associated with REP sequences, RAYTs (REP-associated tyrosine transposases) [14]. Related to transposases of the IS200/IS605 insertion sequence family [21,22], RAYTs carry conserved residues to perform DNA cleavage – the catalytic tyrosine and two metal-coordinating histidines. Since REP elements were found flanking RAYT genes in almost all species where they have been previously recorded, REPs were the likely substrates to be cleaved by RAYTs. The predicted REP-specific nuclease activity of *E. coli* RAYT was recently confirmed experimentally [23], and the crystal structure of REP/RAYT complex was solved [24]. The structure helped to elucidate the role of conserved tetranucleotide and palindrome (two defining features of REP elements) in REP recognition by RAYTs.

Owing to rapid expansion of Next-generation DNA sequencing methods, increasing numbers of new genomic sequences are reported each year. These provide great opportunity to conduct comparative analyses. We explored the distribution of REP elements and their associated RAYTs in sequenced genomes of sixty-three fluorescent pseudomonads and ten stenotrophomonads, two groups of omnipresent environmental bacteria with biotechnological and biocontrol applications [12,25]. Our results indicate rapid diversification and proliferation of REPs in both studied groups. Furthermore, RAYTs appear to play

a principal role in REP dissemination, as RAYT presence correlates with REP abundance. Our results provide support for the hypothesis that REP/RAYT system is an example of mobile element domestication.

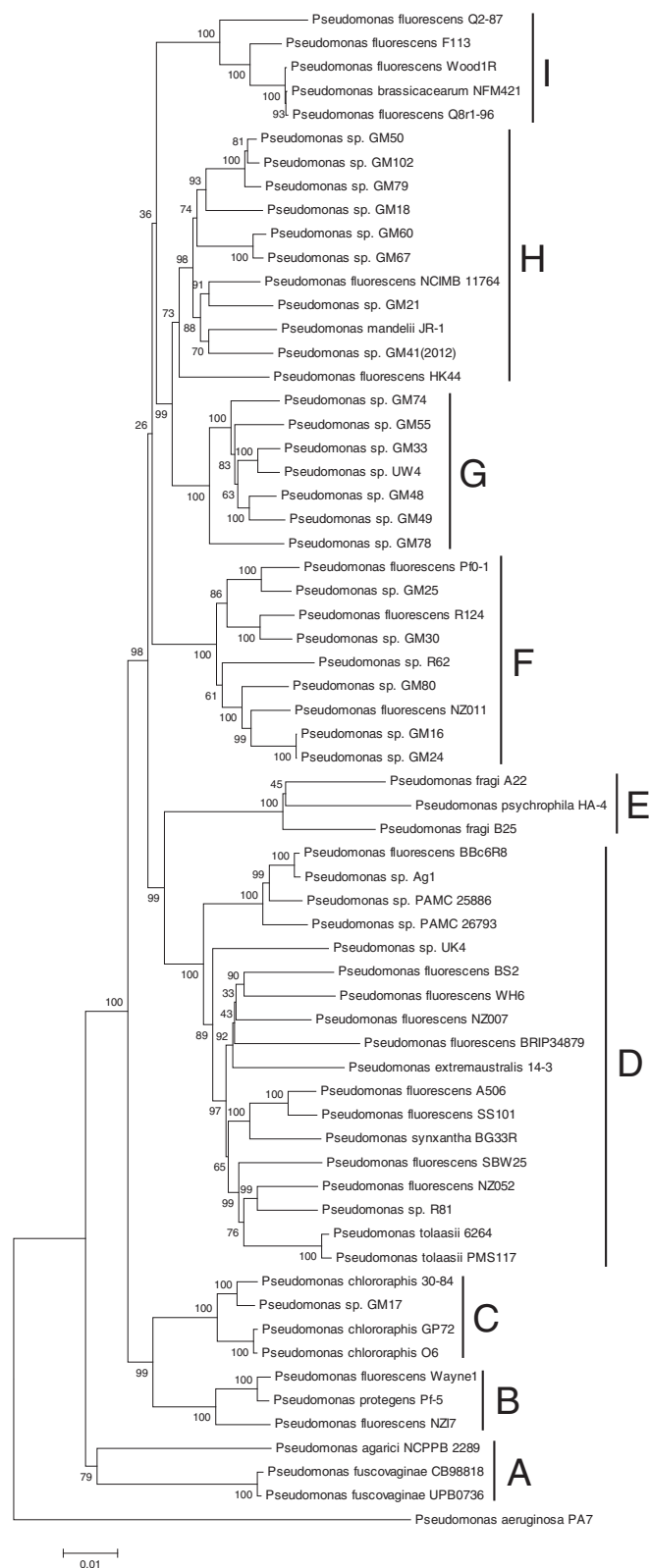
## Results and discussion

### Phylogenetic relationships of studied bacteria

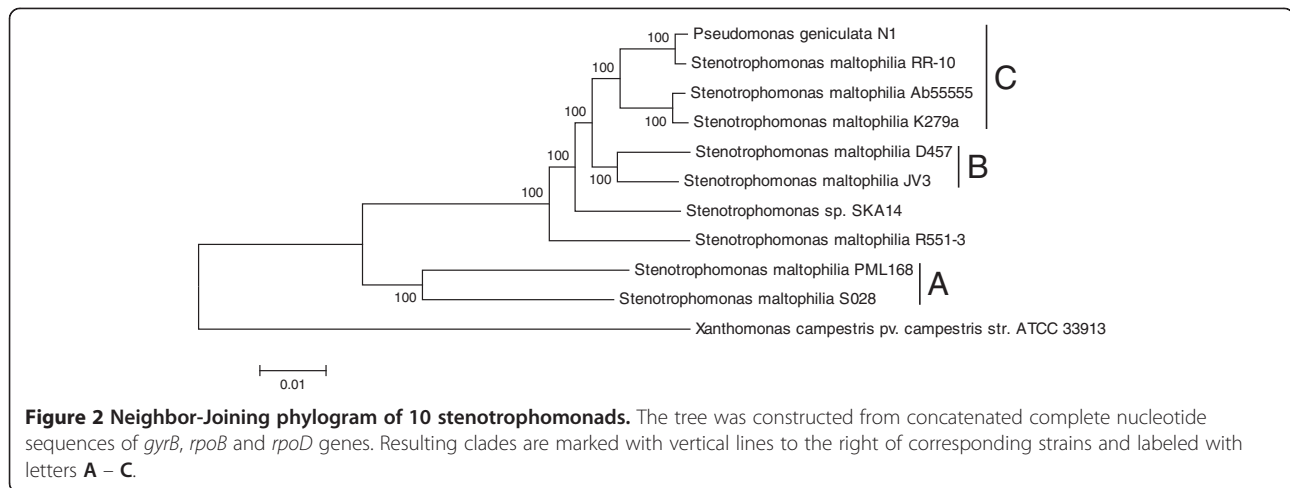
Our preliminary analysis of available genomes revealed that the greatest intraspecific diversity of REP elements and their associated RAYTs existed in bacteria of the *Pseudomonas fluorescens* complex and in *Stenotrophomonas* sp. (data not shown). Comprehensive mining of bacterial genomic databases recovered 63 genomes affiliated to *Pseudomonas fluorescens* (fluorescent pseudomonads) and 10 genomes affiliated to *Stenotrophomonas maltophilia* (stenotrophomonads). Among fluorescent pseudomonads, species of *P. agarici*, *P. brassicacearum*, *P. chlororaphis*, *P. extremaustralis*, *P. fragi*, *P. fuscovaginae*, *P. mandelii*, *P. protegens*, *P. psychrophila*, *P. synxantha* and *P. tolaasii*, previously shown to belong to the *P. fluorescens* complex [26], were included, as well as numerous *Pseudomonas* sp. isolates, unassigned to any species. For stenotrophomonads, *Pseudomonas geniculata*, synonym for *S. maltophilia* [27], was included, as well as *Stenotrophomonas* sp. SKA14. To resolve the evolutionary relationship between the strains, phylogenetic trees were constructed from three housekeeping genes (Figure 1, Figure 2). The phylogram of fluorescent pseudomonads revealed nine well-supported clades (A – I). The phylogram of stenotrophomonads identified three clades (A – C) and two solitary strains. The inter- and intra-clade phylogram resolution was perfect for stenotrophomonads while only partially satisfactory for fluorescent pseudomonads. This difference might be due to the effect of recombination, since *P. fluorescens* was shown to be naturally competent for transformation [28], whereas natural competence is unknown in *S. maltophilia*.

### Diversity of REP sequences and RAYTs

In the next step, the spectrum of REP elements was determined in genomes of studied strains. For this purpose, we utilized the specific association between RAYT (REP-associated tyrosine transposase) genes and REP elements. This approach (see Methods) yielded twenty-two and thirteen unique classes of REP elements in fluorescent pseudomonads (PF1 – PF22) and stenotrophomonads (SM1 – SM13), respectively (Table 1, Table 2, Additional file 1) For some REP classes, sequence ambiguities were detected when two slightly different REP sequences were associated with the same *rayt* gene. REPs of stenotrophomonads always contain eight or nine perfectly complementary bases, located directly adjacent to the GT<sup>A</sup>/<sub>C</sub>G tetranucleotide. In contrast, in fluorescent pseudomonads REPs, palindromes are flanked by additional two or three nucleotides on both sides and the length of the



**Figure 1 Neighbor-Joining phylogram of 63 fluorescent pseudomonads.** The tree was constructed from concatenated complete nucleotide sequences of *gyrB*, *rpoB* and *rpoD* genes. Resulting clades are marked with vertical lines to the right of corresponding strains and labeled with letters **A – I**.



palindromes is significantly shorter (Table 1, Table 2). The majority of detected REPs occurred as close inverted doublets (REPINs), as reported previously [13,15]. The cognate RAYTs of both bacterial groups are monophyletic (Additional file 2), suggesting that although quite diverse,

they have been present in their host genomes for substantial evolutionary time. Intriguingly, several different classes of REP sequences were found to flank orthologous RAYT genes (as judged by their shared chromosomal location - synteny) between related strains in both bacterial sets.

**Table 1 Summary information on identified RAYTs and their cognate REP elements in sequenced fluorescent pseudomonads**

	RAYT/REP symbol	RAYT accession number	Cognate REP sequence
Orthogroup I	PF1	YP_002873491 ( <i>P. fluorescens</i> SBW25)	<b><i>GTGG</i></b> <u>GAGGGGGCTT</u> <b><i>GCCCCGAT</i></b>
	PF2	n.a.	<b><i>GTGG</i></b> <u>GAGGGGGCTT</u> <b><i>GCTCCCGAT</i></b>
Orthogroup II	PF3	n.a.	<b><i>GTAG</i></b> <u>GAGCyGGCTT</u> <b><i>GCCrGCGAA</i></b>
	PF4	EJM82571 ( <i>P. sp.</i> GM60)	<b><i>GTAG</i></b> <u>GAGCCGGCTT</u> <b><i>GCTGGCGAT</i></b>
Orthogroup III	PF5	EJN28792 ( <i>P. sp.</i> GM80)	<b><i>GTGGy</i></b> <u>GAGGGGATT</u> <b><i>TATCCCCG</i></b>
	PF6	n.a.	<b><i>GTGG</i></b> <u>CGAGGGGGCTT</u> <b><i>GTCCCCCG</i></b>
	PF7	EJM60273 ( <i>P. sp.</i> GM49)	<b><i>GTGG</i></b> <u>CGAGGGGGCTT</u> <b><i>GCCCCCG</i></b>
Orthogroup IV	PF8	EIK66912 ( <i>P. fluorescens</i> Q8r1-96)	<b><i>GTGG</i></b> <u>GAGCGAGCTT</u> <b><i>GCTCGCGAT</i></b>
	PF9	EKA23398 ( <i>P. fluorescens</i> BBc6R8)	<b><i>GTGG</i></b> <u>GAGCGGGCTT</u> <b><i>GCTCGCGAA</i></b>
	PF10	EJM47370 ( <i>P. sp.</i> GM33)	<b><i>GTGG</i></b> <u>GAGCGAGCyT</u> <b><i>GCTCGCGAA</i></b>
	PF11	n.a.	<b><i>GTAG</i></b> <u>GAGTGAGCCT</u> <b><i>GCTCGCGAT</i></b>
	PF12	YP_006323329 ( <i>P. fluorescens</i> A506)	<b><i>GTGG</i></b> <u>GAGCTGGCTT</u> <b><i>GCCTGCGAT</i></b>
	PF13	n.a.	<b><i>GTGG</i></b> <u>GAGCGGGCTT</u> <b><i>GCCCCGCGAT</i></b>
	PF14	ZP_10622153 ( <i>P. sp.</i> GM78)	<b><i>GTGG</i></b> <u>GAGCTGGCTT</u> <b><i>GCCAGCGAT</i></b>
	PF15	EJM57603 ( <i>P. sp.</i> GM41 (2012))	<b><i>GTGG</i></b> <u>GAGCCAGCCT</u> <b><i>GCTGGCGAT</i></b>
	PF16	EJM16763 ( <i>P. sp.</i> GM21)	<b><i>GTGG</i></b> <u>GAGCTAGCCT</u> <b><i>GCTAGCGAT</i></b>
	NO*	PF17	YP_002871781 ( <i>P. fluorescens</i> SBW25)
NO*	PF18	ZP_10436910 ( <i>P. extremaustralis</i> 14-3)	<b><i>GTAG</i></b> <u>GAGCGAGCyGCT</u> <b><i>CGCGA</i></b>
NO*	PF19	YP_004351241 ( <i>P. brassicacearum</i> NFM421)	<b><i>GTrG</i></b> <u>GAGCAAGGCTT</u> <b><i>GCCCCGCGAT</i></b>
NO*	PF20	EJM39110 ( <i>P. sp.</i> GM33)	<b><i>GTAG</i></b> <u>GAGCTGCCGAAGGCT</u> <b><i>GCGAT</i></b>
NO*	PF21	EIM18788 ( <i>P. chlororaphis</i> O6)	<b><i>GTAG</i></b> <u>GAGCGAGCCTT</u> <b><i>GCCCCGCGA</i></b>
NO*	PF22	YP_002873800 ( <i>P. fluorescens</i> SBW25)	<b><i>GTrG</i></b> <u>TGAGCGGGCTT</u> <b><i>GCCCCGCGCT</i></b>

REP sequences are denoted in 5' to 3' orientation as follows: conserved tetranucleotide in bold and italics, complementary (palindromic) nucleotides underlined, variable nucleotides (IUPAC code) in lower case.

\* NO – no orthologous RAYT genes flanked by differing REPs were detected.

n. a. – protein not annotated (see Additional file 1 for these sequences).

**Table 2 Summary information on identified RAYTs and their cognate REP elements in sequenced stenotrophomonads**

	RAYT/REP symbol	RAYT accession number	Cognate REP sequence
Orthogroup I	SM1	YP_001970973 ( <i>S. maltophilia</i> K279a)	<b>G</b> <u><i>G</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>G</i></u> <b>G</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>A</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>T</i></u> <b>G</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>G</b> <u><i>G</i></u> <b>C</b> <u><i>A</i></u> <b>C</b>
	SM2	YP_002708831 ( <i>S. sp.</i> SKA14)	<b>G</b> <u><i>G</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>G</i></u> <b>G</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>C</b> <u><i>C</i></u> <b>T</b> <u><i>T</i></u> <b>G</b> <u><i>G</i></u> <b>T</b> <u><i>T</i></u> <b>G</b> <u><i>G</i></u> <b>C</b> <u><i>A</i></u> <b>C</b>
	SM3	YP_006183766 ( <i>S. maltophilia</i> D457)	<b>G</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>w</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>C</b> <u><i>C</i></u> <b>T</b> <u><i>T</i></u> <b>G</b> <u><i>G</i></u> <b>T</b> <u><i>T</i></u> <b>G</b> <u><i>G</i></u> <b>C</b> <u><i>A</i></u> <b>C</b>
Orthogroup II	SM4	YP_002706198 ( <i>S. sp.</i> SKA14)	<b>G</b> <u><i>T</i></u> <b>r</b> <u><i>G</i></u> <b>A</b> <u><i>T</i></u> <b>C</b> <u><i>C</i></u> <b>A</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>T</b> <u><i>G</i></u> <b>C</b> <u><i>G</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>A</i></u> <b>T</b>
	SM5	n.a.	<b>G</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>A</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>G</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>T</b>
Orthogroup III	SM6	n.a.	<b>G</b> <u><i>G</i></u> <b>T</b> <u><i>A</i></u> <b>G</b> <u><i>A</i></u> <b>G</b> <u><i>T</i></u> <b>C</b> <u><i>G</i></u> <b>A</b> <u><i>C</i></u> <b>T</b> <u><i>G</i></u> <b>T</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>T</b> <u><i>C</i></u> <b>G</b> <u><i>A</i></u> <b>C</b> <u><i>T</i></u>
	SM7	n.a.	<b>G</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>m</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>G</i></u> <b>G</b> <u><i>y</i></u> <b>T</b> <u><i>C</i></u> <b>r</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>G</i></u> <b>C</b> <u><i>k</i></u>
NO*	SM8	YP_001972572 ( <i>S. maltophilia</i> K279a)	<b>G</b> <u><i>G</i></u> <b>T</b> <u><i>A</i></u> <b>G</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>C</i></u> <b>T</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>G</i></u> <b>C</b> <u><i>A</i></u> <b>C</b>
NO*	SM9	YP_002030358 ( <i>S. maltophilia</i> R551-3)	<b>T</b> <u><i>G</i></u> <b>T</b> <u><i>A</i></u> <b>G</b> <u><i>A</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>A</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>T</b> <u><i>G</i></u> <b>C</b> <u><i>T</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>T</i></u> <b>C</b>
NO*	SM10	YP_002029847 ( <i>S. maltophilia</i> R551-3)	<b>G</b> <u><i>G</i></u> <b>T</b> <u><i>A</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>T</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>C</b> <u><i>G</i></u> <b>C</b>
NO*	SM11	YP_004793143 ( <i>S. maltophilia</i> JV3)	<b>T</b> <u><i>G</i></u> <b>T</b> <u><i>A</i></u> <b>G</b> <u><i>A</i></u> <b>G</b> <u><i>T</i></u> <b>C</b> <u><i>G</i></u> <b>A</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>A</b> <u><i>T</i></u> <b>G</b> <u><i>C</i></u> <b>T</b> <u><i>C</i></u> <b>G</b> <u><i>A</i></u> <b>C</b> <u><i>T</i></u>
NO*	SM12	n.a.	<b>G</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>A</b> <u><i>G</i></u> <b>T</b> <u><i>C</i></u> <b>G</b> <u><i>A</i></u> <b>C</b> <u><i>T</i></u> <b>T</b> <u><i>G</i></u> <b>C</b> <u><i>T</i></u> <b>C</b> <u><i>G</i></u> <b>A</b> <u><i>C</i></u> <b>T</b>
NO*	SM13	n.a.	<b>G</b> <u><i>T</i></u> <b>A</b> <u><i>G</i></u> <b>A</b> <u><i>G</i></u> <b>C</b> <u><i>C</i></u> <b>C</b> <u><i>A</i></u> <b>C</b> <u><i>C</i></u> <b>G</b> <u><i>T</i></u> <b>T</b> <u><i>G</i></u> <b>T</b> <u><i>C</i></u> <b>C</b> <u><i>G</i></u> <b>G</b> <u><i>C</i></u> <b>T</b>

REP sequences are denoted in 5' to 3' orientation as follows: conserved tetranucleotide in bold and italics, complementary (palindromic) nucleotides underlined, variable nucleotides (IUPAC code) in lower case.

\* NO – no orthologous RAYT genes flanked by differing REPs were detected.

n. a. – protein not annotated (see Additional file 1 for these sequences).

These cases were gathered into so called orthogroups. An orthogroup comprises the classes of REP elements associated with syntenic (orthologous) RAYTs. Three orthogroups were detected in stenotrophomonads and four in fluorescent pseudomonads (Table 1, Table 2), of which orthogroup IV is the most numerous and includes nine distinct REP classes (PF8 - PF16).

#### Variability of REP copy numbers

The copy numbers of particular REP element classes were determined and compared in genomes of related bacterial strains. Table 3 and Table 4 reveal a strikingly uneven distribution of REP sequences among different hosts. High REP abundance was found to be restricted to single strain (PF1 and PF22 in *P. fluorescens* SBW25, SM13 in *S. maltophilia* PML 168), single clade (PF3 in clade B, PF4 in clade H) or several clades (PF8 in clades G and I, PF21 in clades C and H). Various other patterns in distribution can also be detected. REP numbers typically reach hundreds of occurrences of particular REP classes, and are typically more numerous in fluorescent pseudomonads. Here, in four cases, REP numbers exceed a thousand of copies per genome (PF9 in *P. sp.* GM48, *P. sp.* GM79 and *P. fluorescens* R124 and PF10 in *P. fluorescens* NZ011). Typically, several REP classes occur in a single host strain.

#### RAYTs and REP abundance

Finally, we examined if the presence of RAYTs influenced REP abundance. In most cases, RAYTs associated with abundant REP classes were indeed present in host bacterial strains (Table 3, Table 4, Additional file 3, Additional file 4). On average, two to three RAYTs were

present per strain. A maximum of four RAYTs were detected in a single host genome, and several strains contained no RAYTs at all. Sometimes, the RAYT genes contained frameshift or nonsense mutations, indicative of recent pseudogenization. Interestingly, three strains (*P. fluorescens* R124, *P. sp.* UW4 and *P. sp.* GM78) contained two RAYTs associated with two different REPs belonging to the same orthogroup IV. In these cases, one RAYT gene is always located at a novel chromosomal site. This indicates different evolutionary origins of these RAYTs/REPs, for example RAYT duplication followed by mutation of flanking REPs into another REP class of orthogroup IV, or horizontal acquisition and integration of RAYT gene into a novel locus.

The instances when particular REPs were overrepresented while their cognate RAYTs were absent appeared quite often. However, for a great majority of these cases, one of the following was also observed: i) related strains possessed RAYTs associated with REP sequences in question, or ii) RAYTs in given strain were associated with different REP classes, belonging to the same orthogroup (Additional file 3, Additional file 4). As for i), this might indicate loss of RAYT genes from host strain. As for ii), this was represented for example by fluorescent pseudomonads of clade D which harboured REP classes PF9, PF12 and PF13 of orthogroup IV. While multiple copies of each of these REP classes were present, RAYT associated with only one class was detected in each genome. From this, it can be inferred that original REP sequences flanking the RAYT genes have undergone mutations into another REP variants and were subsequently multiplied, leading to the presence of both classes from the same REP orthogroup in

**Table 3 The abundances of 22 REP classes in genomes of 63 sequenced fluorescent pseudomonads**

Bacterial strain	Clade	REP copy number																					
		Ortho group I*		Ortho group II*		Ortho group III*			Ortho group IV*							NO*	NO*	NO*	NO*	NO*	NO*		
		PF 1	PF 2	PF 3	PF 4	PF 5	PF 6	PF 7	PF 8	PF 9	PF 10	PF 11	PF 12	PF 13	PF 14	PF 15	PF 16	PF 17	PF 18	PF 19	PF 20	PF 21	PF 22
<i>P. agarici</i> NCPPB 2289	A	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<i>P. fuscovaginae</i> CB98818		0	0	0	7	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0
<i>P. fuscovaginae</i> UPB0736		0	0	0	6	0	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0	2	0
<i>P. fluorescens</i> NZ17	B	0	0	<u>319</u>	0	0	0	0	1	0	13	4	3	1	0	0	0	<u>40</u>	46	0	0	0	0
<i>P. fluorescens</i> Wayne1		0	0	<u>420</u>	0	0	0	0	0	0	0	0	0	0	0	0	0	4	35	0	0	0	0
<b><i>P. protegens</i> Pf-5</b>		0	0	<u>457</u>	0	0	0	0	0	0	0	0	0	0	0	0	0	1	37	0	0	0	0
<i>P. chlororaphis</i> GP72	C	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	<u>282</u>	0	0	<u>258</u>	0
<i>P. chlororaphis</i> O6		0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	269	0	0	<u>255</u>	0
<i>P. chlororaphis</i> 30-84		0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	<u>297</u>	1	0	17	0
<i>P. sp.</i> GM17		0	0	1	0	0	0	0	4	0	3	0	0	0	0	0	0	0	<u>198</u>	3	0	194	0
<i>P. fluorescens</i> BBc6R8	D	0	0	3	0	0	3	1	0	<u>739</u>	3	0	110	5	0	0	0	155	<u>67</u>	0	0	0	0
<i>P. sp.</i> Ag1		0	0	4	0	0	<u>13</u>	0	0	<u>787</u>	0	0	96	4	0	0	0	154	<u>66</u>	0	0	0	0
<i>P. sp.</i> PAMC 26793		0	0	0	0	0	1	0	0	<u>684</u>	2	0	106	4	0	0	0	155	76	0	0	0	0
<i>P. sp.</i> PAMC 25886		0	0	8	1	0	72	0	0	<u>32</u>	0	0	425	14	3	0	0	<u>117</u>	20	0	0	0	39
<b><i>P. fluorescens</i> A506</b>		15	0	0	0	0	0	0	0	103	0	0	<u>681</u>	63	2	0	0	6	62	0	0	0	0
<i>P. fluorescens</i> SS101		12	1	0	0	0	0	0	0	101	0	0	<u>627</u>	47	1	0	0	3	434	0	0	0	0
<i>P. synxantha</i> BG33R		73	225	0	0	0	0	0	0	15	0	0	<u>326</u>	18	0	0	0	5	64	0	0	0	0
<i>P. fluorescens</i> NZ007		0	0	0	0	0	0	0	1	<u>22</u>	0	0	606	70	0	0	0	<u>54</u>	0	0	0	0	0
<b><i>P. fluorescens</i> SBW25</b>		<u>387</u>	123	0	0	0	0	0	0	6	0	0	43	6	2	0	0	<u>104</u>	0	0	0	0	<u>202</u>
<i>P. sp.</i> R81		45	2	3	0	0	0	0	0	<u>304</u>	0	0	140	9	2	0	0	0	27	0	0	0	0
<i>P. fluorescens</i> NZ052		24	2	11	0	0	1	0	0	199	0	2	226	<u>21</u>	1	0	0	28	<u>281</u>	0	0	0	3
<i>P. tolaasii</i> 6264		2	0	0	0	0	0	0	0	10	1	0	797	<u>151</u>	0	0	0	20	0	0	0	0	0
<i>P. tolaasii</i> PMS117		4	0	0	0	0	0	0	0	10	8	0	824	143	1	0	0	33	0	0	0	0	0
<i>P. fluorescens</i> BRIP34879		12	62	0	0	0	0	0	0	21	0	0	<u>144</u>	21	0	0	0	0	4	0	0	0	0
<i>P. extremaustralis</i> 14-3		20	<u>168</u>	4	1	0	0	0	0	4	0	0	0	2	0	0	0	3	<u>201</u>	0	0	0	0
<i>P. fluorescens</i> BS2		55	<u>181</u>	0	0	0	0	0	0	10	0	0	1	0	0	0	0	1	0	0	0	0	0
<i>P. fluorescens</i> WH6		9	41	0	0	0	0	1	2	4	0	0	2	0	0	0	0	27	0	0	0	0	17
<i>P. sp.</i> UK4		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0

**Table 3 The abundances of 22 REP classes in genomes of 63 sequenced fluorescent pseudomonads (Continued)**

<i>P. psychrophila</i> HA-4	<b>E</b>	0	0	0	0	0	0	0	1	2	0	0	0	4	1	0	0	0	21	0	9	0	0
<i>P. fragi</i> A22		0	0	0	0	0	0	0	0	0	0	0	1	17	116	0	0	0	155	0	0	0	0
<i>P. fragi</i> B25		0	0	0	0	0	0	0	2	17	4	0	67	163	2	0	0	0	0	0	0	0	0
<b><i>P. fluorescens</i> Pf0-1</b>	<b>F</b>	0	0	0	0	0	0	0	3	10	10	0	0	14	31	7	0	0	0	0	0	0	0
<i>P. sp.</i> GM25		0	0	0	0	0	0	0	29	27	66	0	0	4	9	24	8	0	0	0	0	0	0
<i>P. sp.</i> R62		0	0	0	0	<u>150</u>	0	0	0	5	<u>832</u>	7	0	43	145	0	2	0	6	0	99	0	0
<i>P. sp.</i> GM30		0	0	0	0	<u>139</u>	0	0	3	51	<u>582</u>	249	0	19	178	7	0	380	13	0	97	0	0
<i>P. fluorescens</i> R124		0	0	0	0	37	0	0	1	<u>1009</u>	<u>265</u>	217	0	1	30	5	2	<u>261</u>	12	0	<u>59</u>	0	0
<i>P. fluorescens</i> NZ011		0	0	0	0	0	0	0	2	2	<u>1035</u>	240	0	0	6	4	2	2	12	0	111	0	0
<i>P. sp.</i> GM16		0	0	0	0	0	0	0	0	0	159	272	1	14	<u>357</u>	2	8	2	9	0	<u>323</u>	0	0
<i>P. sp.</i> GM24		0	0	0	0	0	0	0	0	0	153	261	1	14	<u>325</u>	2	8	2	9	0	<u>304</u>	0	0
<i>P. sp.</i> GM80		0	0	0	0	<u>306</u>	0	0	0	1	9	182	0	88	<u>535</u>	2	0	4	11	2	0	0	0
<b><i>P. sp.</i> UW4</b>	<b>G</b>	0	0	0	0	0	0	84	<u>427</u>	363	<u>435</u>	0	14	60	48	0	0	2	<u>398</u>	0	25	0	0
<i>P. sp.</i> GM33		0	0	0	1	0	0	92	438	540	<u>156</u>	0	7	34	21	0	0	<u>58</u>	<u>272</u>	0	<u>145</u>	0	0
<i>P. sp.</i> GM48		0	0	1	1	0	0	64	48	<u>1283</u>	8	0	1	5	0	0	0	<u>212</u>	<u>233</u>	0	1	0	0
<i>P. sp.</i> GM49		0	0	1	0	0	0	<u>108</u>	<u>550</u>	151	47	1	4	79	3	0	0	3	<u>502</u>	0	2	0	0
<i>P. sp.</i> GM55		0	0	<u>391</u>	90	0	0	1	33	15	45	0	4	<u>203</u>	35	0	0	1	<u>435</u>	0	0	1	0
<i>P. sp.</i> GM74		0	0	1	0	0	0	0	232	428	11	0	2	120	<u>55</u>	0	0	19	50	0	19	0	0
<i>P. sp.</i> GM78		0	0	17	0	0	0	0	9	12	101	92	1	188	<u>114</u>	<u>117</u>	7	0	150	12	70	53	0
<i>P. fluorescens</i> NCIMB 11764	<b>H</b>	0	0	2	2	51	0	1	2	13	11	0	1	11	5	59	31	<u>186</u>	2	0	12	2	0
<i>P. mandelii</i> JR-1		0	0	0	1	2	0	2	6	5	166	<u>84</u>	0	5	2	3	8	164	68	0	0	<u>355</u>	0
<i>P. fluorescens</i> HK44		0	0	0	0	0	0	56	10	5	29	0	1	0	0	0	0	<u>304</u>	1	100	1	1	0
<i>P. sp.</i> GM50		0	0	0	0	3	0	<u>392</u>	1	<u>198</u>	32	0	4	20	0	15	11	13	4	0	67	<u>667</u>	0
<i>P. sp.</i> GM102		0	0	0	0	0	0	140	2	77	6	0	16	<u>116</u>	1	11	14	<u>245</u>	0	0	57	<u>679</u>	0
<i>P. sp.</i> GM79		0	0	1	400	0	0	16	2	<u>1417</u>	8	0	9	62	0	0	0	127	0	0	168	75	0
<i>P. sp.</i> GM60		0	0	31	<u>398</u>	0	0	0	1	<u>175</u>	30	50	0	5	3	17	6	0	297	0	1	<u>379</u>	0
<i>P. sp.</i> GM67		0	0	20	<u>301</u>	0	0	0	2	33	41	23	0	4	1	37	26	0	101	1	0	<u>656</u>	0
<i>P. sp.</i> GM21		0	0	3	0	1	0	0	12	48	147	20	4	35	4	38	<u>9</u>	0	11	0	0	0	0
<i>P. sp.</i> GM18		0	0	0	0	0	0	27	1	2	3	0	1	7	0	0	0	28	1	1	65	35	0
<i>P. sp.</i> GM41(2012)		0	0	0	1	55	0	73	7	72	91	13	2	13	0	<u>19</u>	5	0	5	<u>174</u>	89	24	0



**Table 3 The abundances of 22 REP classes in genomes of 63 sequenced fluorescent pseudomonads (Continued)**

<i>P. fluorescens</i> Q2-87	<b>I</b>	0	0	0	0	47	0	0	75	<b><u>576</u></b>	130	0	0	2	0	0	0	<b><u>586</u></b>	2	1	0	0	0
<b><i>P. fluorescens</i> F113</b>		0	0	0	0	<b><u>331</u></b>	0	0	<b><u>749</u></b>	60	91	0	0	0	0	0	0	54	3	198	9	0	0
<i>P. fluorescens</i> Q8r1-96		0	0	0	0	30	0	0	<b><u>661</u></b>	61	109	0	0	0	0	0	0	44	5	<b><u>295</u></b>	17	0	0
<i>P. fluorescens</i> Wood1R		0	0	0	0	21	0	0	<b><u>290</u></b>	26	62	0	0	1	0	0	0	37	6	181	15	0	0
<b><i>P. brassicacearum</i> NFM421</b>		0	0	0	0	23	0	0	<b><u>632</u></b>	60	116	0	0	1	0	0	0	46	6	<b><u>303</u></b>	14	0	0

The values represent total numbers of REP sequences from Table 1 in different strains' genomes. The values are denoted in bold and underlined in cases where RAYT gene associated with a particular REP class is present in the genome. Presence of pseudogenized RAYT genes, containing nonsense or frameshift mutations or deletions, is denoted by italics. The phylogenetic groups are marked with letters A to G as in Figure 1. The names of strains whose complete genomic sequences were determined are in bold.

\* - as in Table 1.



**Table 4 The abundances of 13 REP classes in genomes of 10 sequenced stentrophomonads**

Bacterial strain	Clade	REP copy number												
		Ortho group I*			Ortho group II*		Ortho group III*		NO*	NO*	NO*	NO*	NO*	NO*
		SM	SM	SM	SM	SM	SM	SM	SM	SM	SM	SM	SM	SM
		1	2	3	4	5	6	7	8	9	10	11	12	13
<i>S. maltophilia</i> PML168	A	1	0	0	5	<b>18</b>	0	0	37	0	2	1	0	<b>96</b>
<i>S. maltophilia</i> S028		0	0	0	0	0	<b>45</b>	3	0	4	0	0	4	
<b><i>S. maltophilia</i> R551-3</b>	no	<b>39</b>	4	16	62	1	6	0	266	<b>49</b>	<b>259</b>	49	18	0
<i>S. sp.</i> SKA-14	no	7	<b>37</b>	12	<b>128</b>	1	0	0	<b>323</b>	3	7	31	<b>82</b>	2
<b><i>S. maltophilia</i> D457</b>	B	31	8	<b>18</b>	37	2	3	0	<b>258</b>	<b>92</b>	5	15	0	2
<b><i>S. maltophilia</i> JV3</b>		<b>57</b>	4	7	<b>183</b>	2	1	0	<b>283</b>	9	10	<b>108</b>	1	1
<i>S. maltophilia</i> RR-10	C	18	<b>8</b>	10	98	0	6	0	47	<b>106</b>	<b>120</b>	15	2	0
<i>P. geniculata</i> N1		<b>33</b>	7	10	99	1	<b>12</b>	0	61	<b>116</b>	<b>107</b>	18	1	0
<b><i>S. maltophilia</i> K279a</b>		<b>52</b>	16	11	105	2	33	0	<b>427</b>	7	13	3	2	1
<i>S. maltophilia</i> Ab55555		<b>55</b>	13	9	<b>102</b>	4	<b>31</b>	0	<b>375</b>	6	12	3	1	1

The values represent total numbers of REP sequences from Table 2 in different strains' genomes. The values are denoted in bold and underlined in cases where RAYT gene associated with a particular REP class is present in the genome. Presence of pseudogenized RAYT genes, containing nonsense or frameshift mutations or deletions, is denoted by italics. The phylogenetic groups are marked with letters A to C as in Figure 2. The names of strains whose complete genomic sequences were determined are in bold.

\* - as in Table 2.

host genomes. We will call this process an orthoswitch. Although the assumed orthoswitches occurred considerably frequently (*i.e.* at least once in every orthogroup, Table 1 and Table 2), we can only speculate about their molecular mechanism.

In Additional file 5, a model to explain the discrepancies between REP abundance and RAYT presence/absence is proposed. The model assumes an active role of RAYTs in REP proliferation, based on their REP-dependent nuclease activity [23] and coupling of transcription and translation in uncompartimentalized bacterial cell, allowing for preferential RAYT action on REPs that flank their encoding genes (due to their juxtaposition during RAYT expression). According to the model, only the presence of an active RAYT can support multiplication of its cognate REPs and their long-term persistence. When RAYT is inactivated by pseudogenization or completely lost from the host genome, the cognate REPs could no longer multiply, leading to their gradual degradation by mutational processes (Additional file 5A). Depending on when RAYT loss/inactivation occurred, corresponding numbers of REP elements would remain in the host chromosome. Similarly, if an orthoswitch occurred, novel REP variants associated with RAYT genes would spread, while the original REP elements would remain in the host genome and decay mutationally (Additional file 5C). Furthermore, RAYT duplication with concomitant REP diversification (which could proceed with mechanism similar to orthoswitch, see above) would lead to emergence of novel REP classes (Additional file 5B). Finally, horizontal transfer from closely or more distantly related strains might have significantly impacted the REP/RAYT diversity within the

analyzed genomes. Horizontal transfer is likely to have accounted at least for the isolated occurrences of some RAYTs and their cognate REPs (for example PF3 in *P. sp.* GM55, see Table 3).

## Conclusions

In the last decade, there has been a considerable resurgence of interest in REP elements. This was prompted by several factors, notably genomic analyses of newly sequenced bacteria which revealed novel REP elements [29], and the discovery of candidate REP mobilizers, RAYTs [14]. In this study, we aimed to assess the diversity of REP elements and RAYTs in large genomic sets of environmental bacteria – fluorescent pseudomonads and stentrophomonads. Two previous works have already focused on the intraspecific variability of REPs [12,13], but their authors used different, less stringent criteria for REP selection leading to a more relaxed definition of REP classes. We analyzed precisely those REP elements for which association with RAYTs was detected. In addition, our dataset was much broader than those of the two aforementioned studies [12,13]. Our results confirm that REPs of fluorescent pseudomonads and stentrophomonads are very diverse and dynamic. Also, REP host specificity ranges greatly: strain-specific, clade ("subspecies")-specific and species-specific REP sequences were observed (Table 3, Table 4).

Such large-scale analysis of diverse bacteria allowed us to reconstruct the evolutionary scenario for these repetitive elements and associated RAYTs. Since RAYTs of both bacterial groups are monophyletic (Additional file 2), unique original RAYTs were likely to be present in the

genomes of common ancestors of fluorescent pseudomonads and stenotrophomonads, their genes flanked by ancestral REPs. Later during evolution, RAYT genes have undergone duplications and diversified to the state which is seen in more derived clades (Table 3, Table 4), with concomitant diversification of their cognate REPs. The later the novel RAYT/REP variants emerged, the more phylogenetically restricted incidence they show. Novel REP variants might also partially replace the original ones following an orthoswitch (Additional file 5). Upon RAYT pseudogenization which may be followed by RAYT loss from the host genome, proliferation of cognate REPs would cease. Although beneficial roles for REPs have been proposed (see Background), extremely high REP numbers might pose a burden to bacterial hosts, and RAYT inactivation could help keep REP numbers within range tolerable by host cell. A minority of derived strains would lose all RAYTs, leading to greatly reduced REP numbers in their genomes (Table 3).

Since the mechanisms behind REP dissemination and changeability are not known yet, our findings could provide foundations for understanding the evolution of REP element diversity and suggest possible directions for further laboratory research.

## Methods

### Genomic analyses

Bacterial genomic sequences were downloaded from the NCBI Genome database [30]. RAYTs were identified by performing TBLASTN search [31], using previously described *Pseudomonas fluorescens* and *Stenotrophomonas maltophilia* RAYTs [14] as query sequences. RAYTs that were not annotated were conceptually translated from corresponding DNA sequences using Transeq [32]. Identified RAYTs were checked to contain previously characterized sequence motifs peculiar to RAYTs [14]. REP sequences flanking *rayt* genes were identified as inverted repeats located both upstream and downstream of the genes, with characteristic REP features: conserved 5'-terminal tetranucleotide (GT<sup>A</sup>/C<sup>G</sup>) and downstream palindrome. REP copy numbers were determined using pDRAW32 [33].

### Phylogenetic analyses

Concatenated complete nucleotide sequences of genes coding for RNA polymerase beta subunit (*rpoB*), DNA gyrase beta subunit (*gyrB*) and RNA polymerase sigma subunit (*rpoD*) as well as RAYT protein sequences were processed with MEGA5 package [34]. Sequences were aligned, trimmed of unaligned nucleotides or amino acids, and Neighbor-Joining phylograms were constructed with 1 000 bootstrap replicates.

## Additional files

**Additional file 1: Nucleotide sequences of unannotated *rayt* genes and amino acid sequences of translated RAYTs.** The RAYT symbols are as in Table 1 and Table 2.

**Additional file 2: Neighbor-Joining phylogram of representative selection of RAYT proteins across the domain *Bacteria*.** RAYTs of fluorescent pseudomonads and stenotrophomonads are denoted by their symbols as in Table 1 and Table 2. The remaining RAYTs are denoted by their accession numbers and host strain names. All included RAYTs share unique motifs in their sequences and are flanked by inverted repeats with characteristics of REPs (conserved 5'-terminal tetranucleotide and downstream palindromic region). Red circles denote the origins of lineages specific for fluorescent pseudomonads and stenotrophomonads, respectively.

**Additional file 3: Sorting of REP copy numbers in fluorescent pseudomonads based on their association with RAYTs.**

**Additional file 4: Sorting of REP copy numbers in stenotrophomonads based on their association with RAYTs.**

**Additional file 5: A model of long-term evolution of REP elements in genomes of fluorescent pseudomonads and stenotrophomonads.**

The proposed events are: **A** – recent (left) and ancient (right) RAYT loss, **B** – recent (left) and ancient (right) RAYT duplication and diversification, **C** – recent (left) and ancient (right) REP orthoswitch. REP elements (dots) and their associated *rayt* genes (arrows) are denoted with the same color. The host chromosomes are represented as black circles. The REP numbers are approximate. The model is based on two premises: i) RAYTs specifically disseminate their cognate REP elements throughout their host genome, ii) REP copies are getting slowly but constantly lost from host genome, due to accumulation of mutations. See Results and Discussion for further explanation.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JN conceived the study, performed the analysis and drafted the manuscript. BS and IL supervised the work and critically read the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

JN a BS were supported from grant P305/12/1801 from the Czech Science Foundation and by the institutional grant AV0Z50520701. JN and IL were supported by the Charles University grant SVV-2013-267205.

Received: 13 February 2013 Accepted: 3 June 2013

Published: 10 June 2013

### References

1. Koonin EV, Wolf YI: Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* 2008, **36**(21):6688–6719.
2. Mahillon J, Chandler M: Insertion sequences. *Microbiol Mol Biol Rev* 1998, **62**(3):725–774.
3. Redfield RJ, Findlay WA, Bosse J, Kroll JS, Cameron AD, Nash JH: Evolution of competence and DNA uptake specificity in the *Pasteurellaceae*. *BMC Evol Biol* 2006, **6**:82.
4. El Karoui M, Biauudet V, Schbath S, Gruss A: Characteristics of Chi distribution on different bacterial genomes. *Res Microbiol* 1999, **150**(9–10):579–587.
5. Barrangou R, Horvath P: CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 2012, **3**:143–162.
6. Delihans N: Small mobile sequences in bacteria display diverse structure/function motifs. *Mol Microbiol* 2008, **67**(3):475–481.
7. Delihans N: Impact of small repeat sequences on bacterial genome evolution. *Genome Biol Evol* 2011, **3**:959–973.
8. Higgins CF, Ames GF, Barnes WM, Clement JM, Hofnung M: A novel intergenic regulatory element of prokaryotic operons. *Nature* 1982, **298**(5876):760–762.

9. Gilson E, Bachelier S, Perrin S, Perrin D, Grimont PA, Grimont F, Hofnung M: **Palindromic unit highly repetitive DNA sequences exhibit species specificity within *Enterobacteriaceae***. *Res Microbiol* 1990, **141**(9):1103–1116.
10. Aranda-Olmedo I, Tobes R, Manzanera M, Ramos JL, Marques S: **Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida***. *Nucleic Acids Res* 2002, **30**(8):1826–1833.
11. Silby MW, Cerdeno-Tarraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, et al: **Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens***. *Genome Biol* 2009, **10**(5):R51.
12. Loper JE, Hassan KA, Mavrodi DV, Davis EW II, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, et al: **Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions**. *PLoS Genet* 2012, **8**(7):e1002784.
13. Rocco F, De Gregorio E, Di Nocera PP: **A giant family of short palindromic sequences in *Stenotrophomonas maltophilia***. *FEMS Microbiol Lett* 2010, **308**(2):185–192.
14. Nunvar J, Huckova T, Licha I: **Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes**. *BMC Genomics* 2010, **11**(1):44.
15. Bertels F, Rainey PB: **Within-genome evolution of REPINS: a new family of miniature mobile DNA in bacteria**. *PLoS Genet* 2011, **7**(6):e1002132.
16. Gilson E, Saurin W, Perrin D, Bachelier S, Hofnung M: **Palindromic units are part of a new bacterial interspersed mosaic element (BIME)**. *Nucleic Acids Res* 1991, **19**(7):1375–1383.
17. Oppenheim AB, Rudd KE, Mendelson I, Teff D: **Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli***. *Mol Microbiol* 1993, **10**(1):113–122.
18. Espeli O, Boccard F: **In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site**. *Mol Microbiol* 1997, **26**(4):767–777.
19. Gilson E, Perrin D, Hofnung M: **DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization**. *Nucleic Acids Res* 1990, **18**(13):3941–3952.
20. Espeli O, Moulin L, Boccard F: **Transcription attenuation associated with bacterial repetitive extragenic BIME elements**. *J Mol Biol* 2001, **314**(3):375–386.
21. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F: **Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection**. *Cell* 2008, **132**(2):208–220.
22. He S, Guynet C, Siguier P, Hickman AB, Dyda F, Chandler M, Ton-Hoang B: **IS200/IS605 family single-strand transposition: mechanism of IS608 strand transfer**. *Nucleic Acids Res* 2013, **41**(5):13–3302.
23. Ton-Hoang B, Siguier P, Quentin Y, Onillon S, Marty B, Fichant G, Chandler M: **Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences**. *Nucleic Acids Res* 2012, **40**(8):3596–3609.
24. Messing SA, Ton-Hoang B, Hickman AB, McCubbin AJ, Peaslee GF, Ghirlando R, Chandler M, Dyda F: **The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease**. *Nucleic Acids Res* 2012, **40**(19):9964–9979.
25. Ryan RP, Monchy S, Cardinale M, Taghavi S, Crossman L, Avison MB, Berg G, van der Lelie D, Dow JM: **The versatility and adaptation of bacteria from the genus *Stenotrophomonas***. *Nat Rev Microbiol* 2009, **7**(7):514–525.
26. Mulet M, Lalucat J, Garcia-Valdes E: **DNA sequence-based analysis of the *Pseudomonas* species**. *Environ Microbiol* 2010, **12**(6):1513–1530.
27. Svensson-Stadler LA, Mihaylova SA, Moore ER: ***Stenotrophomonas* interspecies differentiation and identification by *gyrB* sequence analysis**. *FEMS Microbiol Lett* 2012, **327**(1):15–24.
28. Demaneche S, Kay E, Gourbiere F, Simonet P: **Natural transformation of *Pseudomonas fluorescens* and *Agrobacterium tumefaciens* in soil**. *Appl Environ Microbiol* 2001, **67**(6):2617–2621.
29. Tobes R, Ramos JL: **REP code: defining bacterial identity in extragenic space**. *Environ Microbiol* 2005, **7**(2):225–228.
30. *NCBI Genome*. <http://www.ncbi.nlm.nih.gov/genome>.
31. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**(17):3389–3402.
32. Rice P, Longden I, Bleasby A: **EMBOSS: The European molecular biology open software suite**. *Trends Genet* 2000, **16**(6):276–277.
33. *pDRAW32 DNA analysis software*. <http://www.acaclone.com/>.
34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Mol Biol Evol* 2011, **28**(10):2731–2739.

doi:10.1186/1471-2164-14-385

**Cite this article as:** Nunvar et al.: Evolution of REP diversity: a comparative study. *BMC Genomics* 2013 **14**:385.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

