Electronic Journal of Biotechnology 34 (2018) 67–75

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research article

Genomic comparisons of *Rhizobium* species using *in silico* AFLP-PCR, endonuclease restriction, and AMPylating enzymes

M. Amjad Qureshi [a,*], Muhammad. Tariq Pervez [a,*], Masroor Ellahi Babar [b], Tanveer Hussain [b], Muhammad Shoaib [c], Syed Shah Mohammad [d]

[a] Department of Bioinformatics and Computational Biology, Virtual University of Pakistan, M.A Jinnah Campus, Defence Road, Off Raiwind Road, Lahore, Pakistan
[b] Department of Molecular Biology, Virtual University of Pakistan, M.A Jinnah Campus, Defence Road, Off Raiwind Road, Lahore, Pakistan
[c] Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Pakistan
[d] Department of Computer Science, Virtual University of Pakistan, M.A Jinnah Campus, Defence Road, Off Raiwind Road, Lahore, Pakistan

ARTICLE INFO

Article history:
Received 3 December 2017
Accepted 22 May 2018
Available online 26 May 2018

Keywords:
AFLP-PCR
AMPylating enzymes
Fabaceae
In silico
Phylogenetic tree
Post-translational modifications
Restriction endonucleases
Restriction enzymes
Rhizobium proteome
Rhizobium species
Rhizobium–legume symbiosis

ABSTRACT

Background: The whole-genome sequences of nine *Rhizobium* species were evaluated using different *in silico* molecular techniques such as AFLP-PCR, restriction digest, and AMPylating enzymes. The entire genome sequences were aligned with progressiveMauve and visualized by reconstructing phylogenetic tree using NTSYS pc 2.11X. The "insilico.ehu.es" was used to carry out *in silico* AFLP-PCR and *in silico* restriction digest of the selected genomes. Post-translational modification (PTM) and AMPylating enzyme diversity between the proteome of *Rhizobium* species were determined by novPTMenzy.
Results: Slight variations were observed in the phylogeny based on AFLP-PCR and PFGE and the tree based on whole genome. Results clearly demonstrated the presence of PTMs, i.e., AMPylation with the GS-ATasE (GlnE), Hydroxylation, Sulfation with their domain, and Deamidation with their specific domains (AMPylating enzymes) GS-ATasE (GlnE), Fic, and Doc (Phosphorylation); Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase; Sulfotransferase; and CNF (Cytotoxic Necrotizing Factors), respectively. The results pertaining to PTMs are discussed with regard to functional diversities reported in these species.
Conclusions: The phylogenetic tree based on AFLP-PCR was slightly different from restriction endonuclease- and PFGE-based trees. Different PTMs were observed in the *Rhizobium* species, and the most prevailing type of PTM was AMPylation with the domain GS-ATasE (GlnE). Another type of PTM was also observed, i.e., Hydroxylation and Sulfation, with the domains Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase and Sulfotransferase, respectively. The deamidation type of PTM was present only in *Rhizobium* sp. NGR234.
How to cite: Qureshi MA, Pervez MT, Babar ME, et al. Genomic comparisons of Rhizobium species using in silico AFLP-PCR, endonuclease restrictions and ampylating enzymes. Electron J Biotechnol 2018;34. https://doi.org/10.1016/j.ejbt.2018.05.006.

1. Introduction

Rhizobium–legume symbiosis is the characteristic of the Fabaceae family, and nodules are considered as a specific feature of legumes. Legumes largely fulfill their nitrogen requirements by biological nitrogen fixation (BNF), i.e., a process in which the nitrogenase enzyme complex converts atmospheric N_2 to ammonia through a cascade of reactions between *Rhizobium* species and legumes [1]. The nitrogen-fixing capacity of *Rhizobium* species varies with host legumes and depends upon the *rhizobium* efficacy, soil, and climatic conditions. *Rhizobium* species, in addition to promoting plant growth through nitrogen-fixing capacity, also promote mutualistic interactions with plants, produce hormones, mobilize nutrients, and relieve stresses [2,3,4,5].

The symbiotic relationship between *Rhizobium* and host legumes provides a prototype for mutualistic evolution and improves the agriculture system on sustainable basis. Approximately 12 genera containing 90 species of both groups (α and β) form nodules and fix nitrogen with leguminous plants [6]. In addition to the presence of remarkable molecular, functional, and genetic resources, i.e., sequenced genes/genomes and protein/amino acid pathways, the molecular basis of natural gene/genome variations between the symbiotic relationships of *Rhizobium* species and legumes was also presented in our study [7].

* Corresponding authors.
E-mail addresses: qureshifsd@gmail.com, (M. Amjad Qureshi), m.tariq@vu.edu.pk (M.T. Pervez).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

https://doi.org/10.1016/j.ejbt.2018.05.006

The ability to differentiate genomic variations at the species or strain level helps in recognizing evolutionary and phylogenetic interactions, functional and morphological variations, and precise taxonomic classifications. The genetic and physiological models of signaling and constant metabolite exchange attracted the evolutionary aspects of *Rhizobium* species owing to the ecological and economic importance of the *Rhizobium*–legume interaction for sustainable agriculture [8,9]. The interactions due to several selective forces resulted in the evolutionary and ecological benefit of symbiosis [10,11,12,13]. *Rhizobium* species are specific to their host, and the genes that are responsible for host specificity are carried in the symbiotic plasmid. In addition to the presence of other phenotypically cryptic plasmids, the plasmid DNA contributes 50% to the genome [14].

Techniques of molecular biology such as amplified fragment length polymorphism-PCR (AFLP-PCR), restriction endonuclease followed by pulsed-field gel electrophoresis (PFGE), and AMPylation enzyme diversity are beneficial for microbial differentiation. AFLP is extensively used as a fingerprinting technique that employs double digestion of genome and fragments; these are then concatenated to specific adaptors, and amplification of some fragments followed by electrophoresis of fragments (amplicons) and visualization of banding patterns is carried out [15].

Restriction enzymes or restriction endonucleases are responsible for cutting DNA at a specific nucleotide. Restriction endonucleases do not usually reveal remarkable differences but can be proven as a useful technique when followed by PFGE [16,17]. AMPylation or adenylylation is a novel post-translational modification (PTM; stable and reversible) in which adenosine monophosphate (AMP) is attached covalently to protein side chains [18]. AMPylation is involved in a variety of biological processes to regulate nitrogen metabolism in prokaryotes and signaling pathways in eukaryotes. AMPylation is catalyzed by three different families of enzymes, namely, Fic (filamentation induced by cAMP), DrrA, and GS-ATase (glutamine synthetase adenylyltransferase) [15,18,19,20,21]. This study was designed to compare *Rhizobium* spp. for their differences based on genomic dissimilarities tapped by AFLP-PCR and restriction endonuclease by using *in silico* methods.

2. Materials and methods

The whole-genome sequences of the following *Rhizobium* species were evaluated with different *in silico* molecular techniques such as AFLP-PCR, restriction endonuclease, and AMPylating enzyme diversity for genome comparison

- *Rhizobium etli* CFN 42
- *R. etli* CIAT 652
- *R. etli* bv. mimosa str. Mim 1
- *Rhizobium leguminosarum* bv. trifolii WSM 1325
- *R. leguminosarum* bv. trifolii WSM 2304
- *R. leguminosarum* bv. viciae 3841
- *Rhizobium* sp. IRBG74
- *Rhizobium* sp. NRG234
- *Rhizobium tropici* CIAT 899

The entire genome sequences of the abovementioned *Rhizobium* species were obtained from NCBI (ftp://ncbi.nih.gov/genomes/bacteria/). The genome sequences were aligned using progressiveMauve [22] to visualize the entire genome sequence resemblance between these species and strains and finally to reconstruct a phylogenetic tree by the neighbor-joining method using NTSYS pc 2.11X (Fig. 1).

Some general characteristics of the *Rhizobium* species under study are given in Table 1. The whole-genome data of different *Rhizobium* species were evaluated using tools available at insilico.ehu.es [15]. AMPylation enzyme diversity was determined using the tool provided by the developer of novPTMenzy [18,23]. The results obtained are presented

Fig. 1. Phylogenetic tree of entire genomes of *Rhizobium* species.

in Table 2, Table 3, and Table 4. *In silico* AFLP-PCR was carried out for genome comparisons of *Rhizobium* species. AFLP-PCR allowed quick scan of the entire genome for polymorphism, generated many bands, and detected polymorphism on the restriction sites. Different restriction enzymes chosen in different combinations enhanced the probability of finding useful polymorphisms in the genomes. Different sets of restriction enzymes, namely, *Eco*RI + *Mse*I, *Eco*RI + *Mbo*I, *Eco*RI + *Taq*I, *Eco52*I + *Asu*II, and *Ec1X*I + *Bsp*XI, with different sets of additional selective nucleotides added to primer sequences were used for each species to generate AFLP bands. The bands were scored as either present or absent (such as binary 1 or 0). Binary data reflecting specific AFLP fragments were analyzed by NTSYS pc 2.11X (Fig. 2). In total, 875 entries of Ntedit 1.2 for different combinations of restriction endonuclease and nucleotides for AFLP-PCR and 435 entries of Ntedit 1.2 for restriction digest and PFGE were recorded.

The endonuclease enzymes selected for restriction digest and *in silico* PFGE (*Ahl*I, *Bcu*I, *Spe*I*Kfl*I; *Bst*SNI, *Eco*105I, *Sna*BI; *Mss*I, *Pme*I; *Pac*I, *Smi*I, and *Swa*I) were used for building a matrix of binary scores. NTSYS pc 2.11X was used for building a phylogenetic tree of *Rhizobium* species (Fig. 3).

For AMPylating enzyme diversity, the protein sequences were retrieved from Uniprot, and these sequences were evaluated for PTM and AMPylating enzyme diversity among *Rhizobium species*. *In silico* AMPylating enzyme diversity was analyzed from a freely available open source website (www.nii.ac.in/novptmenzy.html) that has comprehensive genomic data of *Rhizobium* species regarding AMPylating enzymes. The sequences of *Rhizobium* species under study were obtained in the FASTA format.

Table 1
Some General Statistics of *Rhizobium* species genomes under study.

Rhizobium species	GC (%)	Genome length	ORF	Protein	Size (Mb)	Gene
Rhizobium etli CFN 42	61.3	4381608	4035	4067	4.38	4301
Rhizobium etli CIAT 652	61.7	4513324	4343	4215	4.51	4405
Rhizobium etli bv. mimosa str. Mim 1	61.3	4284494	4064	3970	4.28	4175
Rhizobium leguminosarum bv. trifolii WSM 1325	61.1	4767043	4565	4527	4.77	4710
Rhizobium *leguminosarum* bv. trifolii WSM 2304	61.5	4537948	4325	4322	4.54	4472
Rhizobium leguminosarum bv. viciae 3841	61.1	5057142	4694	4791	5.06	4987
Rhizobium sp. IRBG74	59.3	2844565	2882	2738	2.84	2829
Rhizobium sp. NRG234	63	3925702	3630	3629	3.93	3699
Rhizobium tropici CIAT 899	59.9	3837060	3670	3639	3.84	3727

Table 2
Post-translational modifications of *Rhizobium etli* CFN42, *Rhizobium etli* CIAT652, and *Rhizobium etli* bv. mimosae str. Mim1 with their AMPylating enzyme diversity.

Rhizobium species	Accession number	PTM	Domain/family	Potential active site	Exp. verified homolog	Structural homolog
Rhizobium etli CFN42	Q2KDT8	AMPylation	GLNE	D51, D53	A7Y9V0	3K7D
	Q2K8C5	Hydroxylation	Asparagine_hydroxylase		A7Y9V0	2XUM
	Q2K8V4	Sulfation	Sulfotransferase		O77081	
	Q2KD77	AMPylation	GLNE	D114, D116	Q8CK02	3K7D
	Q2K6V2	Hydroxylation	Asparagine_hydroxylase		A7Y9V0	
	Q2KAN5	AMPylation	GLNE	D204, D206	P30870	3K7D
		AMPylation	GLNE	D719, D721	P30870	3K7D
Rhizobium etli CIAT652	B3PZC7	AMPylation	GLNE	D138, D140	Q79VE2	3K7D
	B3PQN6	Hydroxylation	Collagen_prolyl_lysyl_hydroxylase		Q10576	
	B3Q300	Hydroxylation	Collagen_prolyl_lysyl_hydroxylase		P13674	2V5F
	B3PVA6	AMPylation	Doc (Phosphorylation)	H68, F70, D72, G73, N74, K75, R76	Q8E9K5	3K33
	B3PXK7	AMPylation	GLNE	D51, D53	A7Y9V0	3K7D
	B3PU67	AMPylation	GLNE	D204, D206	P30870	3K7D
		AMPylation	GLNE	D719, D721	P30870	3K7D
Rhizobium etli bv. mimosae str. Mim1	S5S646	Sulfation	Sulfotransferase		Q77081	3AP2
	S5SDI3	AMPylation	GLNE	D138, D140	Q8CK02	3K7D
	S5SJ38	Sulfation	Sulfotransferase		O77081	
	S5SJI6	Hydroxylation	Asparagine_hydroxylase		3P3P	
	S5SH15	AMPylation	GLNE	D204, D206	P30870	3K7D
		AMPylation	GLNE	D719, D721	P30870	3K7D
	S5SVV0	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q4KLM6	
	S5SK45	Hydroxylation	Asparagine_hydroxylase			
	S5RQM7	AMPylation	GLNE	D51, D53	A7Y9V0	3K7D

Table 3
Post-translational modification of *Rhizobium leguminosarum* bv. trifolii WSM 1325, *Rhizobium leguminosarum* bv. trifolii WSM 2304, and *Rhizobium leguminosarum* bv. viciae strain 3841.

Rhizobium species	Accession number	PTM	Domain/family	Potential active site	Exp. verified homolog	Structural homolog
Rhizobium leguminosarum bv. trifolii WSM 1325	C6BAX5	AMPylation	Fic	H337, Y339, D341, G342, N343, G344, R345, R348	Q9BVA6	3CUC
	C6AX32	Sulfation	Sulfotransferase		O77081	
	C6B8G5	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q4KLM6	
	C6B3W8	Hydroxylation	Asparagine_hydroxylase			
	C6AVR1	Hydroxylation	Collagen_prolyl_lysyl_hydroxylase		Q20679	
Rhizobium leguminosarum bv. trifolii WSM 2304	B6A375	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q9GZT9	
	B5ZMZ2	AMPylation	GLNE	D138, D140	Q79VE2	3K7D
	B6A193	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q81VL5	2HBU
	B5ZNB7	Sulfation	Sulfotransferase		O77081	
	B5ZWT6	AMPylation	GLNE	D51, D53	A7Y9V0	3K7D
Rhizobium leguminosarum bv. viciae strain 3841	Q1MIV4	AMPylation	Doc (Phosphorylation)	r, F69, D71, G72, N73, K74, R75	Q9BVA6	3K33
	Q1MDT3	Sulfation	Sulfotransferase		Q3EDG5	
	Q1MJC2	AMPylation	GLNE	D204, D206	P30870	3K7D
		AMPylation	GLNE	D719, D721	P30870	3K7D
	Q1M5C5	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q3V1T4	
	Q1MDX6	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q9H6Z9	3OUJ
	Q1M4T5	Hydroxylation	Collagen_HIF_prolyl_hydroxylase		Q4KLM6	
	Q1MMY0	AMPylation	GLNE	D51, D53	A7Y9V0	3K7D
	Q1MMA6	AMPylation	GLNE	D138, D140	Q79VE2	3K7D
	Q1MEM6	Hydroxylation	Asparagine_hydroxylase			

Table 4
Post-translational modification of *Rhizobium* sp. IRBG74, *Rhizobium* sp. NGR234, and *Rhizobium tropici* CIAT 899 with their AMPylating enzyme diversity.

Rhizobium species	Accession number	PTM	Domain/family	Potential active site	Exp. verified homolog	Structural homolog
Rhizobium sp. IRBG74	U4Q5S7	AMPylation	GLNE	D206, D208	P30870	3K7D
		AMPylation	GLNE	D721, D723	P30870	3K7D
	U4PTE2	AMPylation	GLNE	D116, D118	P30870	3K7D
	U4Q678	AMPylation	Doc (phosphorylation)	H68, F70, D72, G73, N74, K75, R76	Q8SWV6	3K33
Rhizobium sp. NGR234	C3KP39	AMPylation	Fic	n	Q8E9K5	3EQX
	C3MGH0	Hydroxylation	Asparagine_hydroxylase			
	P55548	AMPylation	Fic	H134, F136, E138, G139, N140, G141, R142	E6Z0R3	3SHG
	C3MEZ9	AMPylation	GLNE	D136, D138	P30870	3K7D
	C3MIJ8	Deamidation	CNF			
	P55472	Sulfation	Sulfotransferase		O60507	3AP3
Rhizobium tropici CIAT 899	L0LQ76	AMPylation	Fic	g		3CUC
	N6UFZ8	Sulfation	Sulfotransferase		O60507	3AP2
	L0LYD4	Hydroxylation	Collagen_prolyl_lysyl_hydroxylase		Q20679	
	N6U706	Hydroxylation	Collagen_prolyl_lysyl_hydroxylase			
	N6V5K2	AMPylation	GLNE	D138, D140	Q79VE2	3K7D

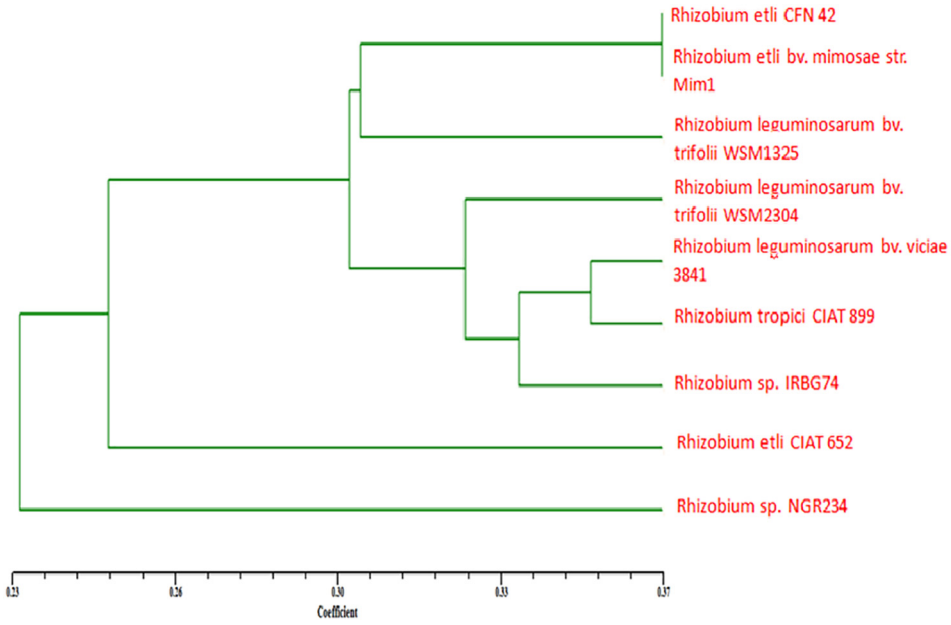

Fig. 2. Phylogenetic tree of *Rhizobium* species developed from AFLP-PCR bands.

3. Results

The genomes of *Rhizobium* species having a variable genome size ranging from 2.84 to 5.06 Mb and gene size ranging from 2829 bp to 4710 bp were used in this study. *R. leguminosarum* bv. viciae 3841 showed the highest genome size, i.e., 5.06 Mb, and *Rhizobium* sp. IRBG74 showed the lowest size, i.e., 2.84 Mb. The GC content ranged from 59.3% to 63%, where *Rhizobium* sp. IRBG74 showed the lowest GC content, i.e., 59.3%, and *Rhizobium* sp. NRG234 exhibited the highest GC content, i.e., 63%. Genome length, open reading frame (ORF), and presence of proteins of *Rhizobium* species have almost similar sizes except *Rhizobium* sp. IRBG74, whose size was 2.84 Mb. Some general characteristics and statistics of *Rhizobium* species genomes are presented in Table 1.

The phylogenetic tree constructed on the basis of whole-genome sequence showed that *R. leguminosarum* spp. like *R. leguminosarum* bv. viciae strain 3841, bv. trifolii WSM 1325, and bv. trifolii WSM2304 clustered together (Fig. 1). However, *R. leguminosarum* bv. viciae strain 3841 and *R. leguminosarum* bv. trifolii WSM1325 showed more resemblance than *R. leguminosarum* bv. trifolii WSM2304. The same was true for *R. etli* spp. where *R. etli* CFN 42 and *R. etli* bv. mimosae strain Mim1 displayed more resemblance to each other than *R. etli* CIAT 652. *R. leguminosarum* spp. clustered together similar to *R. etli* spp., but *R. tropici* spp. related distantly to these clusters. *R. tropici* CIAT 899 related to other species comparatively with wider distance, i.e., *R. leguminosarum* spp. and *R. etli* spp. However, *Rhizobium* sp. NGR234 and *Rhizobium* sp. IRBG74 were the most distinct species from the rest of the species.

The phylogenetic tree on the basis of AFLP-PCR is presented in Fig. 2. *Rhizobium* species were assessed by *in silico* AFLP-PCR available at insilico.ehu.es [15]. Fig. 2 shows that *R. etli* CFN 42 clustered with *R. etli* bv. mimosae str. Mim1, but *R. etli* CIAT 652 showed very a distant relation with other etli species unlike the whole-genome tree. *R. leguminosarum* bv. trifolii WSM 1325 was more closely related to *R. etli* CFN 42 and *R. etli* bv. mimosae str. Mim1 than to *R. leguminosarum* bv. trifolii WSM 2304, *R. leguminosarum* bv. viciae 3841, *R. tropici* CIAT 899, and *R.* sp. IRBG74, which form their own cluster. *R. etli* CIAT 652 was excluded from these clusters, and *Rhizobium* sp. NGR234 was the most distantly related to all species unlike the whole-genome–based tree.

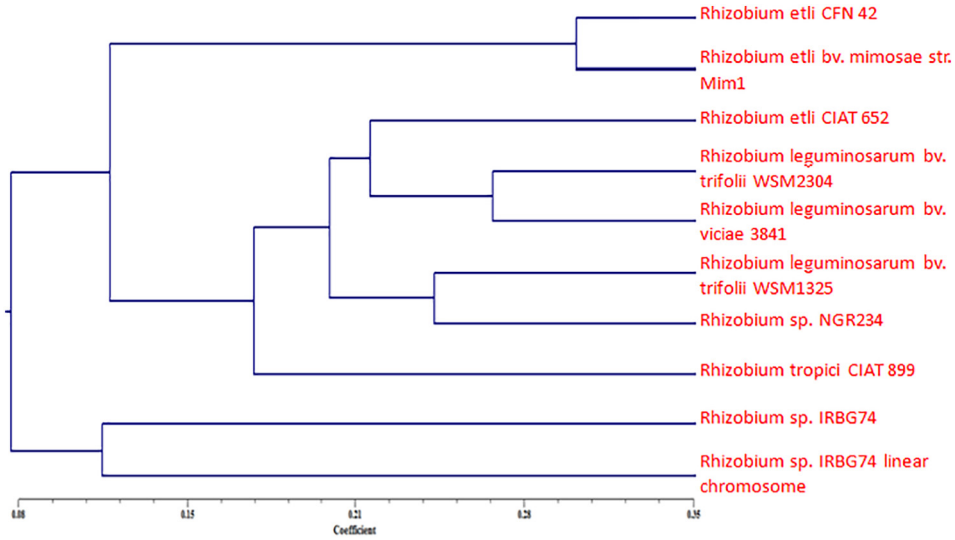

Fig. 3. Phylogenetic tree of *Rhizobium* species developed from endonuclease restriction and PFGE.

Table 5
AFLP-PCR–based similarity matrix among nine *Rhizobium* spp.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	100.00%								
S2	26.27%	100.00%							
S3	36.92%	27.90%	100.00%						
S4	28.02%	20.67%	32.74%	100.00%					
S5	25.00%	24.16%	30.65%	30.70%	100.00%				
S6	32.90%	30.34%	33.15%	33.61%	31.80%	100.00%			
S7	26.13%	20.14%	28.68%	28.38%	33.04%	34.78%	100.00%		
S8	19.72%	19.51%	21.43%	24.39%	25.34%	22.54%	29.57%	100.00%	
S9	28.46%	24.95%	35.56%	28.46%	33.07%	35.38%	32.84%	21.48%	100.00%

S1: *Rhizobium etli* CFN 42; S2: *Rhizobium etli* CIAT 652; S3: *Rhizobium etli* bv. mimosae str. Mim1; S4: *Rhizobium leguminosarum* bv. trifolii WSM1325; S5: *Rhizobium leguminosarum* bv. trifolii WSM2304; S6: *Rhizobium leguminosarum* bv. viciae 3841; S7: Rhizobium sp. IRBG74; S8: Rhizobium sp. NGR234; S9: *Rhizobium tropici* CIAT 89.

The similarity index of *Rhizobium* species is presented in Table 5. *R. etli* CFN 42 was grouped to *R. etli* bv. mimosae str. Mim1 (37% similarity), whereas *R. leguminosarum* bv. viciae 3841 made clade with *R. tropici* CIAT 899 (35.4% similarity). The most distant species, i.e., *Rhizobium* sp. NGR234 from *R. etli* CFN 42, has 19.7% similarity. *Rhizobium* species were investigated using *in silico* tools, i.e., restriction endonuclease and PFGE, available at insilico.ehu.es [15]. *In silico* genomic comparisons of *Rhizobium* species performed by generating a number of bands (435) by different endonucleases (*Ahl*I, *Bcu*I, *Spe*I*Kfl*I, *Bst*SNI, *Eco*105I, *Sna*BI; *Mss*I, *Pme*I; *Pac*I*Smi*, and *Swa*I) were used for phylogenetic analysis (Fig. 3). Fig. 3 shows that *R. etli* CFN 42 and *R. etli* bv. mimosae str. Mim1 were present on the same clade, whereas *R. etli* CIAT 652 related distantly to other etli species. *R. leguminosarum bv*. viciae 3841 and *R. leguminosarum* bv. trifolii WSM 2304 were present on the same clade and *R. leguminosarum* bv. trifolii WSM 1325 was grouped with *Rhizobium* sp. NGR234, whereas relation among members of *R. tropici* CIAT 899 was not close. *Rhizobium* sp. IRBG74 circular as well as linear were present on the same clade.

The similarity index of *Rhizobium* species is presented in Table 6. *R. etli* CFN 42 was grouped to *R. etli* bv. mimosae str. Mim1 (30.22% similarity) and *R. leguminosarum* bv. trifolii WSM2304 and *R. leguminosarum* bv. viciae 3841 made clade with 27% similarity. *R. leguminosarum* bv. trifolii WSM1325 was grouped with *Rhizobium* sp. NGR234 with 23.38% similarity, whereas the least similarity was observed between *R. leguminosarum* bv. trifolii WSM2304 and *Rhizobium* sp. IRBG74, i.e., 0% similarity.

The protein sequences of *Rhizobium* species under study were analyzed using computational tools available at http://www.nii.ac.in/novptmenzy.html. Data regarding PTM of *R. etli* CFN42, *R. etli* CIAT652, and *R. etli* bv. mimosae str. Mim1 with their AMPylating enzyme diversity is presented in Table 2. While evaluating *R. etli* CFN 42, different proteins exhibited PTMs, i.e., AMPylation, Hydroxylation, and Sulfation, with their concerned domains GS-ATasE (GlnE), Asparagine_hydroxylase, and Sulfotransferase, respectively. Results regarding protein sequences of *R. etli* CFN42 with proteins Q2KDT8, Q2KD77, and Q2KAN5 showed AMPylation PTM having domain GS-ATasE (GlnE) with potential active sites (D51, D53), (D114, D116), and (D204, D206), and A7Y9V0, Q8CK02, and P30870 were experimental homolog with 3K7D as the structural homolog. The proteins Q2K8C5 and Q2K6V2 showed Hydroxylation PTM with domain (Asparagine_hydroxylase), and A7Y9V0 and 2XUM were experimental and structural homologs, respectively. The protein Q2K8V4 showed Sulfation PTM with domain Sulfotransferase, and O77081 was the experimental homolog with no structural homolog.

While evaluating *R. etli* CIAT652, different proteins showed PTMs, i.e., AMPylation and Hydroxylation, with their concerned domains GS-ATasE (GlnE) and Collagen_prolyl_lysyl_hydroxylase, respectively. Results regarding the protein sequence of *R. etli* CIAT652, the protein B3PZC7, B3PVA6, B3PXK7, and B3PU67 showed AMPylation PTM having domain GS-ATasE (GlnE), Doc (Phosphorylation) with potential active sites (D138, D140), (H68, F70, D72, G73, N74, K75, R76), (D51, D53), (D204, D206), and (D719, D721). Experimental homologs were Q79VE2, Q8E9K5, and P30870, whereas structural homologs were 3K7D and 3K33. The proteins B3PQN6 and B3Q300 exhibited Hydroxylation PTM with domain (Collagen_prolyl_lysyl_hydroxylase). Their experimental homologs were Q10576 and P13674, and the structural homolog was 2V5F.

While assessing *R. etli* bv. mimosae str. Mim1, the protein sequence showed PTMs, i.e., sulfation, AMPylation, and hydroxylation, with their concerned domains Sulfotransferase; GS-ATasE (GlnE); and Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase, respectively. Results regarding protein sequences of *R. etli* bv. mimosae str. Mim1 showed the proteins S5S646 and S5SJ38 with Sulfation PTM with domain Sulfotransferase whose potential active site was found. The experimental homolog was O77081, while no structural homolog was found. The proteins S5SDI3, S5SH15, and S5RQM7 showed AMPylation PTM with domain (GS-ATasE (GlnE)) with potential active sites (D138, D140), (D204, D206), (D204, D206), and (D51, D53). The experimental homologs were Q8CK02, P30870, and A7Y9V0 and the structural homolog was 3K7D.

While assessing *R. leguminosarum* bv. trifolii WSM 1325, *R. leguminosarum* bv. trifolii WSM 2304 and *R. leguminosarum*

Table 6
Restriction endonuclease/PFGE-based similarity matrix among 10 *Rhizobium* spp.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	100.00%									
S2	27.48%	100.00%								
S3	30.22%	8.82%	100.00%							
S4	2.45%	17.50%	14.29%	100.00%						
S5	21.77%	22.22%	5.26%	18.18%	100.00%					
S6	15.76%	22.22%	9.41%	24.74%	26.97%	100.00%				
S7	11.01%	11.32%	7.02%	2.90%	0.00%	5.71%	100.00%			
S8	12.60%	19.35%	6.06%	10.26%	8.57%	5.06%	11.76%	100.00%		
S9	19.86%	20.29%	8.22%	24.71%	23.38%	19.77%	6.90%	17.91%	100.00%	
S10	0.00%	21.21%	11.43%	19.51%	18.92%	14.46%	3.64%	3.13%	14.08%	100.00%

S1: *Rhizobium etli* CFN 42; S2: *Rhizobium etli* CIAT 652; S3: *Rhizobium etli* bv. mimosae str. Mim1; S4: *Rhizobium leguminosarum* bv. trifolii WSM1325; S5: *Rhizobium leguminosarum* bv. trifolii WSM2304; S6: *Rhizobium leguminosarum* bv. viciae 3841; S7: *Rhizobium* sp. IRBG74; S8: *Rhizobium* sp. IRBG74 linear chromosome; S9: *Rhizobium* sp. NGR234; S10: *Rhizobium tropici* CIAT 899.

bv. viciae strain 3841 (Table 3), the protein sequences showed PTMs, i.e., AMPylation, Sulfation, and Hydroxylation, with their concerned domains Fic, Sulfotransferase, GS-ATasE (GlnE), and Collagen_prolyl_lysyl_hydroxylase, respectively.

For results regarding protein sequences of *R. leguminosarum* bv. trifolii WSM 1325, the protein C6BAX5 showed AMPylation PTM having domain Fic that starts with potential active sites (H337, Y339, D341, G342, N343, G344, R345, R348). The experimentally found homolog was Q9BVA6, and the structural homolog was 3CUC. The protein C6AX32 having Sulfation PTM with domain (Sulfotransferase) has no potential active site, and the experimentally determined homolog was O77081. The proteins C6B8G5, C6B3W8, and C6AVR1 showed Hydroxylation PTM with domain (Collagen_HIF_prolyl_Hydroxylase, Asparagine_hydroxylase, and Collagen_HIF_prolyl_hydroxylase). The experimentally identified homologs were Q4KLM6 and Q20679.

While assessing *R. leguminosarum* bv. trifolii WSM 2304, the protein sequences showed PTMs, i.e., Hydroxylation, AMPylation, and Sulfation, with their concerned domains Collagen_prolyl_lysyl_hydroxylase, GS-ATasE (GlnE), and Sulfotransferase, respectively. Results regarding protein sequences with the proteins B6A375 and B6A193 of *R. leguminosarum* bv. trifolii WSM 2304 showed Hydroxylation PTM with domain (Collagen_HIF_prolyl_Hydroxylase). In this case, no potential active site was identified. The experimentally determined homologs were Q9GZT9 and Q81VL5, whereas the structural homolog was 2HBU. The proteins B5ZMZ2 and B5ZWT6 showed AMPylation PTM having domain (GS-ATasE (GlnE)). The potential active sites were (D138, D140) and (D51, D53), whereas the experimentally found homologs were Q79VE2 and A7Y9V0 and the structural homolog was 3K7D.The protein B5ZNB7 showed Sulfation PTM with domain (Sulfotransferase). No potential active site was observed, and the experimentally identified homolog was O77081.

While assessing *R. leguminosarum* bv. viciae strain 3841, the protein sequences showed PTMs, i.e., AMPylation, Sulfation, and Hydroxylation, with their concerned domains Doc (Phosphorylation) and GS-ATasE (GlnE); Sulfotransferase; and Collagen_prolyl_lysyl_hydroxylase and Asparagine_hydroxylase, respectively. The proteins Q1MIV4, Q1MJC2, Q1MMY0, and Q1MMA6 of *R. leguminosarum* bv. viciae strain 3841 showed AMPylation PTM having domain Doc (Phosphorylation), GS-ATasE (GlnE) with potential active sites (r, F69, D71, G72, N73, K74, R75), (D204, D206), (D719, D721), (D51, D53), and (D138, D140). Their experimental homologs were Q9BVA6, P30870, A7Y9V0, and Q79VE2, and structural homologs were 3K33 and 3K7D. The protein Q1MDT3 showed Sulfation PTM with domain (Sulfotransferase) with no potential active site. The experimentally identified homolog was Q3EDG5. The proteins Q1M5C5, Q1MDX6, Q1M4T5, and Q1MEM6 showed Hydroxylation PTM with domain (Collagen_prolyl_lysyl_hydroxylase, Asparagine_hydroxylase). The experimentally identified homologs were Q3V1T4, Q9H6Z9, and Q4KLM6, and the structural homolog was 3OUJ.

Data regarding PTM of *Rhizobium* sp. IRBG74, *Rhizobium* sp. NGR234, and *R. tropici* CIAT 899 with their AMPylating enzyme diversity is presented in Table 4. While evaluating *Rhizobium* sp. IRBG74, the protein showed AMPylation type of PTM with their concerned domains GS-ATasE (GlnE) and Doc (Phosphorylation), respectively. For results regarding protein sequences of *Rhizobium* sp. IRBG74, the proteins U4Q5S7, U4PTE2, and U4Q678 presented AMPylation PTM having domain GS-ATasE (GlnE) and Doc (Phosphorylation) with potential active sites (D206, D208), (D721, D723), (D116, D118), and (H68, F70, D72, G73, N74, K75, R76). The experimentally determined homologs were P30870 and Q8SWV6, and the structural homologs were 3K7D and 3K33.

While evaluating *Rhizobium* sp. NGR234, the protein sequence showed diverse PTMs, i.e., AMPylation, Hydroxylation, Deamidation, and Sulfation, with their concerned domains Fic and GS-ATasE (GlnE); Asparagine_hydroxylase; and Sulfotransferase, respectively. The proteins C3KP39, P55548, and C3MEZ9 of *Rhizobium* sp. NGR234 showed AMPylation PTM having domain Fic and GS-ATasE (GlnE) with potential active sites (n), (H134, F136, E138, G139, N140, G141, R142) and (D136, D138). The experimentally verified homologs were Q8E9K5, E6Z0R3, and P30870 and their structural homologs were 3EQX, 3SHG, and 3K7D. Results regarding Hydroxylation PTM having domain Asparagine_hydroxylase demonstrated no experimentally verified homolog and structural homolog. The protein C3MIJ8 with Deamidation PTM having domain CNF illustrated no experimentally verified homolog and structural homolog, whereas Sulfation PTM having domain Sulfotransferase showed O60507 as the experimentally verified homolog and 3AP3 as the structural homolog.

While evaluating *R. tropici* CIAT 899, the protein sequences showed PTMs, i.e., AMPylation, Sulfation, Hydroxylation, and Deamidation with their concerned domains Fic, GS-ATasE (GlnE); Sulfotransferase; and Collagen_prolyl_lysyl_hydroxylase, respectively. For results regarding protein sequences of *R. tropici* CIAT 899, the proteins L0LQ76 and N6V5K2 demonstrated AMPylation PTM having domains Fic and GS-ATasE (GlnE), with D138 and D140 as the potential active sites. The experimentally verified homolog was Q79VE2, and the structural homologs were 3CUC and 3K7D. The protein N6UFZ8 revealed Sulfation PTM having domain Sulfotransferase, with O60507 as the experimentally verified homolog and 3AP2 as the structural homolog. The protein "L0LYD4" and "N6U706" (Uncharacterized protein) demonstrated Hydroxylation PTM having domain Collagen_prolyl_lysyl_Hydroxylase. The experimentally verified homolog was Q20679.

4. Discussion

The resultant phylogenetic tree shown in Fig. 1 clearly indicates that *R. etli* species and *R. leguminosarum* species were found to be clustered together, and *Rhizobium* sp. IRBG74, *Rhizobium* sp. NGR 234, and *R. tropici* CIAT 899, however, were placed distantly.

AFLP is an inexpensive, competent, and convenient technique for genomic research and phylogenetic analysis from the last decade, especially where DNA sequences were unavailable [24,25,26]. AFLP involves digestion with restriction endonuclease, ligation of fragments to adapters, annealing of PCR primers to each adapter, and expansion with specific nucleotides used to amplify the fragments that were visualized and organized according to their length by electrophoresis. The results obtained by the AFLP technique were highly reproducible and recorded as the presence or absence of bands and inferred for phylogenetic relationships [25,27].

The results produced by the AFLP technique showed that sequence divergence is short and symmetric [28]. The reliability of the phylogenetic tree was based on AFLP, which also has a negative impact if the number of bands is small. Nevertheless, at least 500–1000 bands are necessary to build a more reliable tree. A range of 300–500 bands was inadequate to build precise and perfect relationships, although some clades might be precisely rebuilt [28].

In silico genomic comparison showed the phylogeny of bacterial species at a particular similarity index, and some particular clades were formed. It was revealed that when phylogeny based on AFLP-PCR was compared with the whole-genome phylogeny, observed variations might be due to the number of bands, nucleotide usage, existence of the problem of homolog-assigned fragments, prevailing asymmetry in the losing and gaining fragments, and differentiation between homozygous or heterozygous bands. The findings are in accordance with those found in [24,28], where it is found that slight variations in the AFLP-PCR–based phylogenetic tree were in reference with the whole-genome tree. The AFLP bands were scored as the presence or absence of the bands, and any loss of band might cause alteration in the restriction site [29]. *R. etli* CFN 42 was found to be clustered with *R. etli* bv. mimosae str. Mim1, and this was exactly similar to that found in the whole-genome tree; this finding might indicate the appropriateness of AFLP bands for phylogenetic reconstruction [25,27].

Another molecular technique, i.e., PFGE, was deemed as the gold standard owing to its remarkable discriminatory power and profound epidemiological grade. Restriction endonucleases and PFGE have been applied to a broad spectrum of microbes and involve digestion with restriction endonucleases, size separation of broken fragments of genomic DNA in the agarose, and fragments disjointing by the PFGE [30,31,32,33]. The bands of each species were compared and counted for the restriction fragment bands that were observed as identical (no difference), closely related, and/or dissimilar. Numerous molecular techniques were used for bacterial typing, but PFGE could weakly relate bacterial species particularly distant species [34,35,36].

In silico genomic comparisons through restriction endonucleases and PFGE showed the phylogenetic relationships of bacterial species at a specific similarity index, and some bacterial species were found to be clustered. Results revealed that when the phylogenetic tree based on restriction endonuclease and PFGE was compared with the whole-genome phylogeny, slight variations were observed, and these variations might be due to the number of bands, false-positive– or false-negative–related results, and differentiation between homozygous or heterozygous bands. The findings are verified by many researchers who observed slight deviations in PFGE-based phylogeny relationships with reference to the whole-genome tree [37, 38]. The restriction endonucleases and PFGE bands were scored, and they attained a position where large fragments or fragments with minor dissimilarity in molecular size were present, and they appeared as a single blurry band by amending the electrical field between pairs of positive-negative electrode formed by isolated bands [38,39]. *R. etli* CFN 42 was grouped with *R. etli* bv. mimosae str. Mim1, which was rightly similar to the whole-genome phylogenetic tree and might be suitable for restriction endonuclease and PFGE bands for the establishment of a phylogenetic tree [39,40,41,42]. The restriction endonuclease and PFGE-based phylogeny showed variations in the relatedness or resemblance of *Rhizobium* species genomes as compared to the whole-genome phylogeny.

Results clearly indicated the presence of PTMs in *Rhizobium* species, i.e., AMPylation, Hydroxylation, Sulfation, and Deamidation, with their specific domains GS-ATasE (GlnE), Fic, and Doc (Phosphorylation); Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase; Sulfotransferase; and CNF (Cytotoxic Necrotizing Factors), respectively.

The sequence was aligned first, and then, the potential active site in that sequence was displayed [18]. The potential active sites displayed the ATP-binding sites that were more likely to be conserved in the GS-ATasE domain [43,44]. Machine learning approaches such as Support Vector Machines (SVM) for the classification of complex dataset and Hidden Markov Models (HMM) were implemented to determine the cataloging and recognition of various protein domains and to assess the categorization of subfamilies. The HMM was also employed to determine the experimentally verified homolog [45,46].

Glutamine synthetase adenylyltransferase (GS-ATase) regulates Glutamine Synthetase (GS) where the C-terminal domain Adenylyl Transferase adenylylation (AMPylation) Glutamine Synthetase and the N-terminal domain Adenylyl Removase deadenylylation (deAMPylation) Glutamine Synthetase are present in most of the bacteria [47]. The protein sequences were searched through the entire database of novPTMenzy using HAMMER tool [48] and certain accession numbers related to their PTMs such as AMPylation, Hydroxylation, Sulfation, and Deamidation with their specific domains, namely, GS-ATasE, Fic, and Doc (Phosphorylation); Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase; Sulfotransferase; and CNF (Cytotoxic Necrotizing Factors), respectively.

The AMPylating enzymes responsible for AMPylation belonged to three distinct families of enzymes, namely, Fic/Doc or Fido, GS-ATase, and DrrA. The three diverse families of enzymes are Filamentation induced by cyclic AMP (Fic), Glutamine Synthetase Adenyltransferases (GSATase), and Defects in Rab1 recruitment protein A (DrrA) [49,50, 51]. The Fic and DrrA domains of AMPylation were involved in the modulation of host proteins by virulent pathogens. The Fic domain was involved in neurotransmission in glial cells; in other eukaryotes, it was observed to result in complications of other biological processes [21,52]. The Fic type of enzymes established in all types of prokaryotes and eukaryotes catalyzed various PTMs to the target proteins in an inhibited form and altered the horizontal gene transfer [47,53].

The Doc (Phosphorylation) was also observed for AMPylation PTM in addition to GS-ATase (GLnE) that was responsible for death on curing (Doc). It was quite similar to Fic proteins, and thus, it was grouped as the Fic/Doc family. The Fic/Doc family, in addition to AMPylation, is responsible for catalyzing phosphorylation and phosphocholine. The Doc family was involved in phosphorylation of AnkX proteins (Anyrin repeat-containing protein X) [54,55,56].

Hydroxylation was another PTM with the domains Asparagine_hydroxylase and Collagen_HIF_prolyl_hydroxylase. It was responsible for hypoxia-inducible transcription factor (HIF) by hydrolysis of proline and asparagine and regulation of different proteins such as ankyrin repeat-containing proteins (AnkX protein) with asparagine and aspartic acid [57,58,59,60]. Results confirmed the findings of many workers, where YcfD of *Escherichia coli* was homologous to the arginine residue of ribosomal protein (Rpl16) and to AvrB of *Pseudomonas syringae* [61,62,63].

Sulfation was another type of PTM having the domain Sulfotransferase for the covalent transfer of a sulfate group to tyrosine and other proteins such as membrane proteins, coagulation factors, protein-related secretions, and immunity [64]; sulfation was also responsible for the modulation of protein interactions and biological activities [64,65, 66].

Another kind of PTM was deamidation, which adversely affected the structure and function of the proteome and their role in enhancing the metabolic and signaling pathways. The deamidation was esteemed by typical processing conditions such as temperature, pH, and salting of hams. The CNFs, namely, CNF1, CNF2, CNF3, and CNFc of *E. coli,* resulted in the deamidation of a glutamine residue to glutamate, thus conclusively iterating the host the Rho GTPases [67,68,69].

The metabolic regulation of protein pathways in different *Rhizobium* species by modulating the GS-ATasE (GlnE), Fic, Doc (Phosphorylation), Asparagine_hydroxylase, Collagen_prolyl_lysyl_Hydroxylase, Sulfotransferase, and CNF was characterized, and experimentally verified homologs were determined from many organisms, namely, *Azospirillum brasilense*, *Streptomyces coelicolor*, *E. coli* K12, *Corynebacterium glutamicum*, *Caenorhabditis elegans*, *Homo sapiens*, *Shewanella oneidensis*, *Rattus norvegicus*, *Mus musculus*, *Drosophila melanogaster*, and *Bartonella schoenbuchensis* [43,44,70].

The structural homologs of bacterial species were observed and characterized, such as 3K7D, 2XUM, 2V5F, 3AP2, 3P3P, 3CUC, 2HBU, 3K33, 3OUJ, 3EQX, and 3SHG, and these homologs were conserved in many prokaryotes and eukaryotes. This finding corroborated with the findings of [70] who found the homologs of 3K7D in *E. coli and Mycobacterium tuberculosis,* [52,71] found the homolog of 3CUC in *Bacteroides thetaiotaomicron,* and [18,23] reported the conserved nature of Fic/Doc and the lack of a conserved motif in AvrB of *Pseudomonas syringae*.

The present study results concluded that the phylogenetic tree based on AFLP-PCR, restriction endonucleases, and PFGE revealed slight variations. The phylogenetic tree based on AFLP-PCR was slightly different from restriction endonuclease- and PFGE-based trees; different clades were formed with these techniques. Slight similarities in the same clades were also observed with the trees formed by these techniques. Different PTMs were observed in the *Rhizobium* species, and the most prevailing type of PTM was AMPylation with the domain GS-ATasE (GlnE). Other types of domain/AMPylating enzyme diversity were Fic and Doc (Phosphorylation). Other types of PTMs were also observed, i.e., Hydroxylation, followed by Sulfation, with the domains Asparagine_hydroxylase and Collagen_prolyl_lysyl_hydroxylase and Sulfotransferase, respectively. The deamidation type of PTM was only

present in *Rhizobium* sp. NGR234 and *R. tropici* CIAT 899. Results also confirmed and concluded that AFLP-PCR and endonucleases restrictions followed by PFGE can successfully be used for the genotypic similarities among species. The AMPylating enzyme diversities can also be used for the functional variations among bacterial species.

References

[1] Bhattacharyya PN, Jha DK. Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. World J Microbiol Biotechnol 2012;28(4):1327–50. https://doi.org/10.1007/s11274-011-0979-9.

[2] Costa PB, Beneduzi A, Souza R, et al. The effects of different fertilization conditions on bacterial plant growth promoting traits: Guidelines for directed bacterial prospection and testing. Plant and Soil 2013;368(1-2):267–80. https://doi.org/10.1007/s11104-012-1513-z.

[3] Granada CE, Arruda L, Lisboa BB, et al. Diversity of native rhizobia isolated in south Brazil and their growth promotion effect on white clover (*Trifolium repens*) and rice (*Oryza sativa*) plants. Biol Fertil Soils 2014;50(1):123–32. https://doi.org/10.1007/s00374-013-0840-4.

[4] Granada CE, Costa PB, Lisboa BB, et al. Comparison among bacterial communities present in arenized and adjacent areas subjected to different soil management regimes. Plant and Soil 2013;373(1-2):339–58. https://doi.org/10.1007/s11104-013-1796-8.

[5] Kempel A, Nater P, Fischer M, et al. Plant microbe-herbivore interactions in invasive and non-invasive alien plant species. Funct Ecol 2013;27(2):498–508. https://doi.org/10.1111/1365-2435.12056.

[6] Masson-Boivin C, Giraud E, Perret X, et al. Establishing nitrogen-fixing symbiosis with legumes: How many *rhizobium* recipes? Trends Microbiol 2009;17(10): 458–66. https://doi.org/10.1016/j.tim.2009.07.004.

[7] Heath KD, Burke PV, Stinchcombe JR. Coevolutionary genetic variation in the legume-*rhizobium transcriptome*. Mol Ecol 2012;21(19):4735–47. https://doi.org/10.1111/j.1365-294X.2012.05629.x.

[8] Ferguson BJ, Indrasumunar A, Hayashi S, et al. Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol 2010;52(1):61–76. https://doi.org/10.1111/j.1744-7909.2010.00899.x.

[9] Riely B, Ané J, Penmetsa R, et al. Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. Curr Opin Plant Biol 2004;7(4):408–13. https://doi.org/10.1016/j.pbi.2004.04.005.

[10] Heath KD, Lau JA. Herbivores alter the fitness benefits of a plant-*rhizobium* mutualism. Acta Oecol 2011;37(2):87–92. https://doi.org/10.1016/j.actao.2010.12.002.

[11] Heath KD, Tiffin P. Stabilizing mechanisms in a legume *rhizobium* mutualism. Evolution 2009;63(3):652–62. https://doi.org/10.1111/j.1558-5646.2008.00582.x.

[12] Oono R, Anderson CG, Denison RF. Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. Proc R Soc Lond Ser B 2011;278(1718):2698–703. https://doi.org/10.1098/rspb.2010.2193.

[13] Sachs JL, Ehinger MO, Simms EL. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. J Evol Biol 2010;23(5):1075–89. https://doi.org/10.1111/j.1420-9101.2010.01980.x.

[14] Prakash RK, Atherly AG. Plasmids of *Rhizobium* and their role in symbiotic nitrogen fixation. Int Rev Cytol 1986;104:1–24. https://doi.org/10.1016/S0074-7696(08)61921-X.

[15] Bikandi J, Millán RS, Rementeria A, et al. In silico analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. Bioinformatics 2004;20 (5):798–9. https://doi.org/10.1093/bioinformatics/btg491.

[16] González-Escalona N, Whitney B, Lee-Ann J, et al. Comparison of direct genome restriction enzyme analysis and pulsed-field gel electrophoresis for typing of *Vibrio vulnificus* and their correspondence with multilocus sequence typing data. Appl Environ Microbiol 2007;3(22):7494–500. https://doi.org/10.1128/AEM.00738-07.

[17] Sabat AJ, Budimir A, Nashev D, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 2013;18(4): 20380. https://doi.org/10.2807/ese.18.04.20380-en.

[18] Khater S, Mohanty D. novPTMenzy: A database for enzymes involved in novel post-translational modifications. Database 2015:bav039. https://doi.org/10.1093/database/bav039.

[19] Ham H, Woolery AR, Tracy C, et al. Unfolded protein response-regulated *Drosophila* Fic (dFic) reversibly AMPylates BiP during endoplasmic reticulum homeostasis. J Biol Chem 2014;289(52):36059–69. https://doi.org/10.1074/jbc.M114.612515.

[20] Itzen A, Blankenfeldt W, Goody RS. Adenylylation: Renaissance of a forgotten post-translational modification. Trends Biochem Sci 2011;36(4):221–8. https://doi.org/10.1016/j.tibs.2010.12.004.

[21] Rahman M, Ham H, Liu X, et al. Visual neurotransmission in *Drosophila* requires expression of Fic in glial capitates projections. Nat Neurosci 2012;15(6):871–5. https://doi.org/10.1038/nn.3102.

[22] Darling ACE, Mau B, Blattner FR, et al. Mauve: Multiple alignment of conserved genomics sequence with rearrangements. Genome Res 2004;14(7):1394–403. https://doi.org/10.1101/gr.2289704.

[23] Khater S, Mohanty D. In silico identification of AMPylating enzymes and study of their divergent evolution. Sci Rep 2015;5(10804). https://doi.org/10.1038/srep10804.

[24] Koopman WJM. Phylogenetic signal in AFLP data sets. Syst Biol 2005;54(2):197–217. https://doi.org/10.1080/10635150590924181.

[25] Meudt HM, Clarke AC. Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci 2007;12(3):106–17. https://doi.org/10.1016/j.tplants.2007.02.001.

[26] Vos P, Hogers R, Bleeker M, et al. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res 1995;23(21):4407–14. https://doi.org/10.1093/nar/23.21.4407.

[27] Althoff DM, Gitzendanner MA, Segraves KA. The utility of amplified fragment length polymorphisms in phylogenetics: A comparison of homology within and between genomes. Syst Biol 2007;56(3):477–84. https://doi.org/10.1080/10635150701427077.

[28] García-Pereira MJ, Caballero A, Quesada H. The relative contribution of band number to phylogenetic accuracy in AFLP data sets. J Evol Biol 2011;24(11): 2346–56. https://doi.org/10.1111/j.1420-9101.2011.02361.x.

[29] Simmons MP, Zhang L, Webb CT, et al. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. Mol Phylogenet Evol 2007;42(2):528–42. https://doi.org/10.1016/j.ympev.2006.08.008.

[30] Tenover FC, Arbeit RD, Goering RV, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: Criteria for bacterial strain typing. J Clin Microbiol 1995;33(9):2233–9.

[31] Harmsen D, Claus H, Witte W, et al. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. J Clin Microbiol 2003;41(12):5442–8. https://doi.org/10.1128/JCM.41.12.5442-5448.2003.

[32] Foxman B, Zhang L, Koopman JS, et al. Choosing an appropriate bacterial typing technique for epidemiologic studies. Epidemiol Perspect Innov 2005;2(10). https://doi.org/10.1186/1742-5573-2-10.

[33] Foley SL, Lynne AM, Nayak R. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram negative bacterial foodborne pathogens. Infect Genet Evol 2009;9(4):430–40. https://doi.org/10.1016/j.meegid.2009.03.004.

[34] Blanc DS, Francioli P, Hauser PM. Poor value of pulsed-field gel electrophoresis to investigate long-term scale epidemiology of methicillin-resistant *Staphylococcus aureus*. Infect Genet Evol 2002;2:145–8. https://doi.org/10.1016/S1567-1348(02)00093-X.

[35] Malachowa N, Sabat A, Gniadkowski M, et al. Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, spa typing, and multi-locus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. J Clin Microbiol 2005;43(7):3095–100. https://doi.org/10.1128/JCM.43.7.3095-3100.2005.

[36] Nemoy LL, Kotetishvili M, Tigno J, et al. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase producing *Escherichia coli* isolates. J Clin Microbiol 2005;43(4):1776–81. https://doi.org/10.1128/JCM.43.4.1776-1781.2005.

[37] Orsi RH, Borowsky ML, Lauer P, et al. Short-term genome evolution of Listeria monocytogenes in a non-controlled environment. BMC Genomics 2008;9(2008): 539. https://doi.org/10.1186/1471-2164-9-539.

[38] Kyoui D, Takahashi H, Miya S, et al. Genetic distance in the whole-genome perspective on *Listeria monocytogenes* strains F2-382 and NIHS-28 that show similar subtyping results. BMC Microbiol 2014;14(309). https://doi.org/10.1186/s12866-014-0309-0.

[39] Salipante SJ, SenGupta DJ, Cummings LA, et al. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J Clin Microbiol 2015;53(4):1072–9. https://doi.org/10.1128/JCM.03385-14.

[40] Doleans-Jordheim A, Cournoyer B, Bergeron E, et al. Reliability of *Pseudomonas aeruginosa* semi-automated rep-PCR genotyping in various epidemiological situations. Eur J Clin Microbiol Infect Dis 2009;28(9):1105–11. https://doi.org/10.1007/s10096-009-0755-z.

[41] Maâtallah M, Bakhrouf A, Habeeb MA, et al. Four genotyping schemes for phylogenetic analysis of *Pseudomonas aeruginosa*: Comparison of their congruence with multi-locus sequence typing. PLoS One 2013; 2015;10(9):e0137931. https://doi.org/10.1371/journal.pone.0082069.

[42] Ratkai C, Peixe LV, Grosso F, et al. Successful application of the Diversi Lab repetitive-sequence-based PCR typing system for confirmation of the circulation of a multi-resistant *Pseudomonas aeruginosa* clone in different hospital wards. Diagn Microbiol Infect Dis 2010;67(2):202–6. https://doi.org/10.1016/j.diagmicrobio.2010.01.010.

[43] Xu Y, Carr PD, Vasudevan SG, et al. Structure of the adenylylation domain of *E. coli* glutamine synthetase adenylyltransferase: Evidence for gene duplication and evolution of a new active site. J Mol Biol 2009;396(3):773–84. https://doi.org/10.1016/j.jmb.2009.12.011.

[44] Xu Y, Zhang R, Joachimiak A, et al. Structure of the N-terminal domain of *Escherichia coli* glutamine synthetase adenylyltransferase. Structure 2004;12(5):861–9. https://doi.org/10.1016/j.str.2004.02.029.

[45] Ansari MZ, Sharma J, Gokhale RS, et al. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinf 2008;9: 454. https://doi.org/10.1186/1471-2105-9-454.

[46] Yadav G, Gokhale RS, Mohanty D. Towards prediction of metabolic products of polyketide synthases: An in-silico analysis. PLoS Comput Biol 2009;5(4):e1000351. https://doi.org/10.1371/journal.pcbi.1000351.

[47] Jiang P, Ninfa AJ. Reconstitution of *Escherichia coli* glutamine synthetase adenylyltransferase from N-terminal and C-terminal fragments of the enzyme. Biochem 2009;48(2):415–23. https://doi.org/10.1021/bi801775b.

[48] Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011;7(10): e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

[49] Anderson WB, Stadtman ER. Glutamine synthetas edeadenylation: A phosphorolytic reaction yielding ADP as nucleotide product. Biochem Biophys Res Commun 1970; 41(3):704–9.

[50] Yarbrough ML, Li Y, Kinch LN, et al. AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. Science 2009;323(5911): 269–72. https://doi.org/10.1126/science.1166382.

[51] Muller MP, Peters H, Blumer J, et al. The Legionella effector protein DrrA AMPylates the membrane traffic regulator Rab1b. Science 2010;329(5994):946–9. https://doi.org/10.1126/science.1192276.
[52] Kinch LN, Yarbrough ML, Orth K, et al. Fido, a novel AMPylation domain common to fic, doc, and AvrB. PLoS One 2009;4(6):e5818. https://doi.org/10.1371/journal.pone.0005818.
[53] Stanger FV, Burmanna BM, Harmsb A, et al. Intrinsic regulation of FIC-domain AMP-transferases by oligomerization and automodification. PNAS 2016;113(5):E529–37. https://doi.org/10.1073/pnas.1516930113.
[54] Anantharaman V, Aravind L. New connections in the prokaryotic toxin–antitoxin network: Relationship with the eukaryotic nonsense-mediated RNA decay system. Genome Biol 2003;4(12):R81. https://doi.org/10.1186/gb-2003-4-12-r81.
[55] Mukherjee S, Liu X, Arasaki K, et al. Modulation of RabGTPase function by a protein phosphocholine transferase. Nature 2011;477(362):103–6. https://doi.org/10.1038/nature10335.
[56] Castro-Roa D, Garcia-Pino A, De Gieter S, et al. The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. Nat Chem Biol 2013;9(12):811–7. https://doi.org/10.1038/nchembio.1364.
[57] Lando D, Peet DJ, Whelan DA, et al. Asparagine hydroxylation of the HIF transactivation domain: A hypoxic switch. Science 2002;295(5556):858–61. https://doi.org/10.1126/science.1068592.
[58] Metzen E. Ratcliffe PJ. HIF hydroxylation and cellular oxygen sensing. Biol Chem 2004;385(3-4):223–30. https://doi.org/10.1515/BC.2004.016.
[59] Yang M, Ge W, Chowdhury R, et al. Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. J Biol Chem 2011;286(9):7648–60. https://doi.org/10.1074/jbc.M110.193540.
[60] Hudson DM, Eyre DR. Collagen prolyl 3-hydroxylation: A major role for a minor post-translational modification? Connect Tissue Res 2013;54(4-5):245–51. https://doi.org/10.3109/03008207.2013.800867.
[61] Ge W, Wolf A, Feng T, et al. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. Nat Chem Biol 2012;8(12):960–2. https://doi.org/10.1038/nchembio.1093.
[62] Chowdhury R, Sekirnik R, Brissett NC, et al. Ribosomal oxygenases are structurally conserved from prokaryotes to humans. Nature 2014;510(7505):422–6. https://doi.org/10.1038/nature13263.
[63] van Staalduinen LM, Novakowski SK, Jia Z. Structure and functional analysis of YcfD, a novel 2-oxoglutarate/Fe(2)(þ)-dependent oxygenase involved in translational regulation in *Escherichia coli*. J Mol Biol 2014;426(9):1898–910. https://doi.org/10.1016/j.jmb.2014.02.008.
[64] Stone MJ, Chuang S, Hou X, et al. Tyrosine sulfation: An increasingly recognised post-translational modification of secreted proteins. N Biotechnol 2009;25(5):299–317. https://doi.org/10.1016/j.nbt.2009.03.011.
[65] Kehoe JW, Bertozzi CR. Tyrosine sulfation: A modulator of extracellular protein–protein interactions. Chem Biol 2000;7(3):R57–61. https://doi.org/10.1016/S1074-5521(00)00093-4.
[66] Woods AS, Wang HY, Jackson SN. Sulfation, the up-and-coming post-translational modification: Its role and mechanism in protein–protein interaction. J Proteome Res 2007;6(3):1176–82. https://doi.org/10.1021/pr060529g.
[67] Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein post-translational modifications: The chemistry of proteome diversifications. Angew Chem 2005;44(45):7342–72. https://doi.org/10.1002/anie.200501023.
[68] Knust Z, Schmidt G. Cytotoxic necrotizing factors (CNFs)—A growing toxin family. Toxins 2010;2(1):116–27. https://doi.org/10.3390/toxins2010116.
[69] Cañete M, Mora L, Toldra F. Deamidation post-translational modification in naturally generated peptides in Spanish dry-cured ham. Food Chem 2017;229:710–5. https://doi.org/10.1016/j.foodchem.2017.02.134.
[70] Nilsson MT, Krajewski WW, Yellagunda S, et al. Structural basis for the inhibition of *Mycobacterium tuberculosis* glutamine synthetase by novel ATP-competitive inhibitors. J Mol Biol 2009;393(2):504–13. https://doi.org/10.1016/j.jmb.2009.08.028.
[71] Woolery AR, Luong P, Broberg CA, et al. AMPylation: Something old is new again. Front Microbiol 2010;1(113). https://doi.org/10.3389/fmicb.2010.00113.