Chapter 10
Hierarchical Clustering-Based Algorithms and In Silico Techniques for Phylogenetic Analysis of Rhizobia

Jyoti Lakhani, Ajay Khuteta, Anupama Choudhary, and Dharmesh Harwani

10.1 Introduction

Evolution can be defined as the development of a species by divergence of it from other pre-existing species. The driving force behind evolution is natural selection in which "unfit" forms are eliminated through changes of environmental conditions or sexual selection so that only the fittest are selected (Darwin 1859). Mutation is the mechanism behind the evolution that occurs spontaneously to provide the biological diversity within a population. The development of bioinformatics tools and various in silico methods has provided very useful and fast methods to perform phylogenetic analysis. Two types of methods are most commanly used for it: distance based and character based. The distance-based methods include unweighted paired group method with arithmetic mean (UPGMA) (Murtagh 1984), minimum evolution method (ME) (Rzhetsky and Nei 1993), neighbour joining (NJ) (Saitou and Nei 1987), and Fitch–Margoliash method (FM) (Fitch and Margoliash 1967). The character-based method derives trees that optimize the distribution of the actual data pattern for each character. The most commonly used character-based methods include Maximum Parsimony (MP) method (Sober 1983) and Maximum Likelihood (ML) method (Felsenstein 1981). The criteria to

J. Lakhani
Department of Computer Science, Poornima University, Jaipur, India

Department of Computer Science, Maharaja Ganga Singh University, Bikaner, India

A. Khuteta
Department of Computer Science, Poornima University, Jaipur, India

A. Choudhary
Department of Computer Science, Keen College, Bikaner, India

D. Harwani (✉)
Department of Microbiology, Maharaja Ganga Singh University, Bikaner, India
e-mail: dharmesh@mgsubikaner.ac.in

A.P. Hansen et al. (eds.), *Rhizobium Biology and Biotechnology*, Soil Biology 50,
DOI 10.1007/978-3-319-64982-5_10

compare different tree-building methods are computational speed, consistency of estimated topology, statistical consistency of phylogenetic trees, probability of obtaining the correct topology, and reliability of estimated branch length (Roy et al. 2014). According to the computational speed, the NJ method is the superior one from other tree-building methods which are currently in use. This method can handle a large number of sequences with bootstrap tests with ease. If no bias is applied during the estimation of distance through substitution NJ, ME methods are found consistent for estimating trees but MP is often inconsistent. ML methods, on the other hand, have the additional advantage of being more flexible in choosing the evolutionary model. But this method is lengthy and time consuming (Roy et al. 2014). This chapter is a compressive survey on phylogenetic analysis of rhizobia at molecular level. The contributions of few authors who have used hierarchical clustering to assess rhizobial phylogeny have been summarized. The chapter is divided into three sections which include the introduction to the basics and process of molecular phylogenetic analysis, a brief discussion on various hierarchical algorithms and finally, a detailed discussion on different in silico phylogenetic analysis tools to study evolution and phylogeny in rhizobia has been presented.

10.2 Molecular Phylogenetic Analysis

Molecular phylogenetic analysis is the study of relationship among organisms using molecular markers such as DNA or protein sequences. The dissimilarity between two sequences has been caused by mutations during the course of time. The methods in molecular phylogenetic analysis make assumptions about the processes of molecular evolution over time and the accuracy of predicted evolutionary events are tested using in silico simulations. The results of these methods are hypothetical evolutionary trees or phylogenetic trees. Phylogenetic trees are dendograms representing evolutionary divergence between two sequences. There are several types of evolutionary trees such as rooted trees also called cladograms, unrooted trees, or phenogram. The process of generation of a hypothetical phylogenetic tree is called phylogenetic reconstruction. Phylogenetic reconstruction is a probability-based statistical model to make assumptions about the process of nucleotide or amino acid substitution during the timeline in question. There are several types of probabilistic models also which are known as evolutionary models. Evolutionary models describe the different probabilities of the change from one nucleotide or amino acid to other, with the aim of correcting for unseen changes along the phylogeny. The most common models of DNA evolution are Jukes–Cantor (JC or JC69) (Jukes and Cantor 1969), Kimura2 Parameters (K2P or K80) (Kimura 1980), Felsenstien (F81) (Felsenstein 1981) and Hasegawa, Kishino, Yano (HKY85) (Hasegawa et al. 1985), T92 (Tamura 1992), TN93 model (Tamura and Nei 1993), GTR: Generalised time-reversible (Tavaré 1986), *etc*. The common amino acid replacement models are point accepted mutation (PAM) (Dayhoff et al. 1978), mtREV, JTT, WAG, BLOSUM62 (BLOck SUbstitution Matrix), Yang, etc. Apart

from evolutionary models, alignment of the sequences is also a prerequisite for phylogenetic tree construction. There are several multiple sequence alignment methods available such as ClustalW, Muscle, and NAST. A phylogenetic tree is constructed using distance matrix by examining the closeness of sequences in order to combine them. There are several methods used in literature for constructing phylogenetic trees such as UPGMA, neighbour-joining, maximum parsimony, maximum likelihood, and Baysian analysis.

10.3 Basics of Phylogeny

A phylogeny is a graphical representation that provides a hypothesis of how organisms are related at evolutionary level. The relationships are not expressed as per cent sequence similarity, but time since they share a common ancestor. Phylogenetic trees are a primary tool used in evolutionary biology and are used to interpret the timing and order of evolutionary events. Charles Darwin has used tree for the first time to represent phylogeny. Figure 10.1 is the only figure in Charles Darwin's book Origin of Species by Natural Selection (1859) depicting evolutionary history. Some modern applications of phylogeny include analysis of changes that have occurred during the evolution in order to create tree of life of for various organisms, phylogenetic relationships among genes predicting similar functions in order to detect orthologues, detecting changes in rapidly changing sequences, etc.

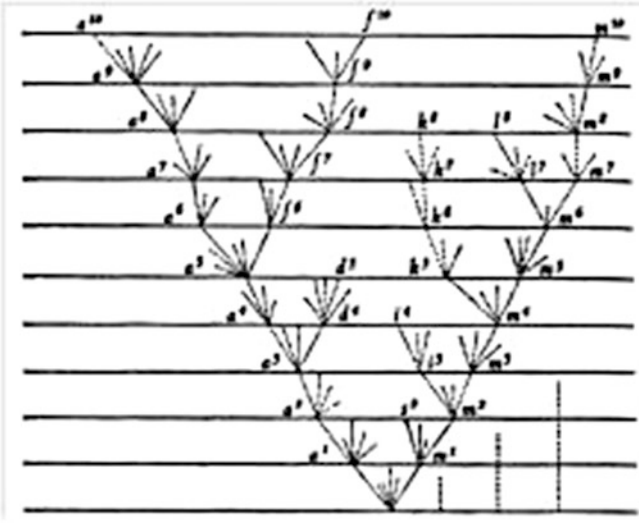

Fig. 10.1 First use of phylogenetic tree to show the evolutionary history of an organism (Origin of Species by Natural Selection 1859)

To display phylogenetic trees, two fundamental forms are used such as rooted trees and unrooted trees. The root of a tree represents the common ancestor of all depicted organisms. All trees need not to be rooted, but rooting does help to interpret tree. Trees are rooted with the inclusion of an outgroup, a taxon known a priori to be the most distant taxon to the group under study. The tips of a tree are referred to as external nodes which typically represent living or extant taxa and ancestors are represented by internal (ancestral) nodes. The phylogenetic topology is the patterns of branch length and splitting depict evolution, diversification, and relatedness. Topology illustrates the history of cladogenesis (splitting of branches as a result of diversification) and anagenesis (change within lineages such as mutation or substitution). In general, diversification events should be dichotomous (one lineage splits into two); however, trees may not be completely dichotomous. Polytomies are common when one computes a consensus tree (a topology that agrees with those found in several trees). These are trees that are generated from bootstrap analysis with many replicates (the fusion of multiple high scoring trees that should be considered as candidates). Lengths illustrate divergence in the characters used to construct the phylogeny (substitutions in DNA sequence). To infer the evolutionary history of an organism, different molecular markers such as DNA, RNA, and protein sequences are used. DNA or protein sequences from homologous (orthologous) genes or proteins from different organisms have been aligned using sequence alignment algorithms. Sequence alignments are arrangements of multiple DNA or protein sequences that tend to minimize the number of gaps and mismatches if an alignment is done judiciously. Hence, sequence alignment is a major tool in construction of a phylogenetic tree. There are three methods for constructing phylogenetic trees: maximum parsimony, distance measure, and maximum likelihood. Maximum parsimony is employed when the evolutionary distances between taxa are relatively short and assumes the rate of mutation among all sequences are equal. Maximum parsimony is based on Fitch's algorithm which is a bottom-up dynamic programming framework for evaluating the parsimony of a given tree and treats each sequence locus as independent of the rest.

Maximum likelihood is often used to construct trees for publication, with the cost of time-consuming processing and is most sensitive when working sequences spanning large evolutionary distances. Maximum likelihood is a robust method that outperforms alternative methods such as parsimony and distance methods (UPGMA) but it is computationally very intensive; therefore, it is slow on most computers. The popular phylogenetic maximum likelihood algorithms are PHYLIP, RAxML, genetic algorithm for rapid likelihood inference (GARLI), PHYML, etc. Statistical support for a phylogenetic tree has performed by a bootstrap analysis. Distance methods are often used to generate a starting tree for the maximum likelihood method and are important to understand the functionality of these three methods in detail in order to construct an approximate real tree of evolution. Distance methods aim to identify the tree that minimizes sequence divergence. The idea behind this approach is that the minimum sequence divergence minimizes evolution. These methods do not utilize an alignment during the tree

search; instead they use a pairwise distance matrix. Distance matrix can be computed by determining the proportion of nucleotides that differ between all pairs.

Distance method is a stepwise process which includes five basic steps—alignment of sequences, computation of pairwise distances between sequences, applying evolutionary correction, construction of tree (Hierarchical Clustering) and evaluating tree, and selecting the best one. There are several sequence alignment tools available such as ClustalW, Muscle, and NAST. The simplest method to find pairwise dissimilarity is Hamming distance which can find number of mismatches. Hamming distance does not take into account the likelihood of one amino acid to other. These problems can be addressed by assigning these sequences a number in order to associate with each possible alignment. The scoring scheme is a set of rules which assigns the alignment score to any given alignment of two sequences. The scoring scheme is residue based: it consists of residue substitution scores, minus penalties for gaps**.** The alignment score is the sum of substitution scores and gap penalties. Point accepted mutation (PAM matrices) and Blocks Substitution Matrix (BLOSUM) are substitution matrices for amino acid alignment. Different versions of PAM and BLOSUM Substitution Matrix are given in Table 10.1 (Source NCBI).

Given the computed distance matrix from above, we could construct a tree. However, how do we know that multiple mutations haven't occurred at the same locus? Multiple substitutions can be caused by enough evolutionary time, high mutation rates, action of positive natural selection. It is quite possible homologous nucleotide positions have undergone multiple substitutions. To generate distance values that correct for multiple hits, one can perform the Jukes–Cantor correction or the Kimura 2-paramter model. Jukes–Cantor correction assumes that all types of mutations/substitutions occur at the same rate. Kimura two-parameter model corrects for multiple hits, giving differential weight to transitions and transversions. In the next step, we can construct tree using hierarchical clustering. UPGMA is the most popular hierarchical clustering algorithm used in the research to construct a single rooted phylogenetic tree. The basic assumption of UPGMA is that distance from any node to leaf will be the same for all common descendants and there is a constant rate of evolution. Two sequences with shortest evolutionary distance between them are assumed to have been the last to diverge. UPGMA is very computationally efficient and provides a good starting point for more sophisticated phylogenetic analysis. However, some issues with UPGMA are that it is very sensitive to unequal evolutionary rates and clustering only works if data is ultrametric (the evolutionary rate is the same for all branches).

Table 10.1 Different versions of PAM and BLOSUM substitution matrix

Query length	Substitution matrix	Gap cost
<35	PAM-30	(9,1)
35–50	PAM-70	(10,1)
50–85	BLOSUM -80	(10,1)
85	BLOSUM-62	(10,1)

Source: https://www.ncbi.nlm.nih.gov/blast/html/sub_matrix.html

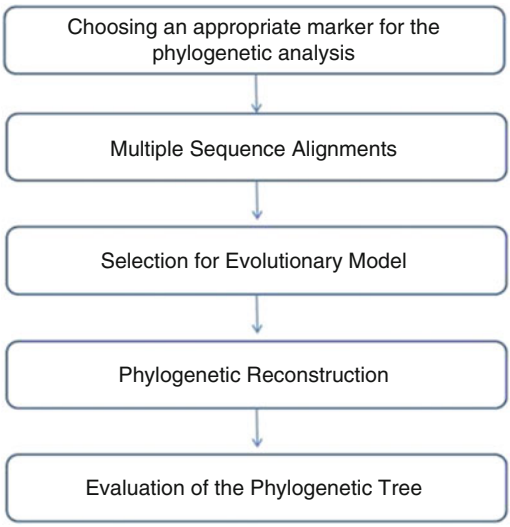

Fig. 10.2 The construction of a phylogenetic tree

10.4 Phylogenetic Tree Construction Using Hierarchical Clustering Algorithms and Tools

When talking about phylogenetic analysis, hierarchical clustering algorithms are unignorable. Given a set of sequences, hierarchical clustering algorithms, cluster these sequences and seek to build a hierarchy of clusters based on the differences. These algorithms work behind the construction of phylogenetic tree (Fig. 10.2).

Two different types of hierarchical algorithms are available in literature—agglomerative and divisive strategies. Agglomerative hierarchical clustering is a bottom-up approach where each sequence is considered as a cluster in its own. These singleton clusters merge with other clusters when one moves up in hierarchy. On the other hand, divisive hierarchical clustering algorithm is a top-down approach in which all sequences start in one cluster and splits are performed as one moves down in hierarchy. The results of both these hierarchical clustering are dendrograms representing phylogenetic trees.

10.5 Hierarchical Clustering Algorithms

UPGMA (Unweighted Pair-Group Method using arithmetic Averages) is probably the most popular hierarchical algorithm for computational biology. D'haeseleer has used UPGMA for gene expression analysis and Liu and Rost have used it for protein sequence clustering. UPGMA was used for gene ontology (GO) by Ashburner et al. and classifies genes into hierarchies of biological processes and molecular functions. ProtoNet was used to build a hierarchy of protein sequences

from sequence similarities. This way UPGMA can be used for a variety of phylogenetic analysis. UPGMA has been used as a phylogenetic tree construction tool for rhizobia number of researches (Blažinkov et al. 2007; Abdel-Aziz et al. 2008; Faisal et al. 2009; Dourado et al. 2009; Jurelevicius et al. 2010; Lyra et al. 2013; Jia et al. 2015; Hassen et al. 2014; Baginsky et al. 2015). The other algorithms for hierarchical clustering that are not very popular such as AGNES, DIANA, BIRCH, ROCK, Chameleon, and CURE but have also been referred in this chapter.

10.6 Hierarchical Clustering Tools

Besides hierarchical algorithms, other hierarchical clustering tools for evolutionary study of rhizobia are also available in literature. R package is a statistical tool having a variety of functions related to sequence analysis (Bontemps et al. 2005; Vercruysse et al. 2011; Knief et al. 2011; Tian et al. 2012; McGinn et al. 2016). Another tool is SPSS that is basically a statistical tool but have some plugins available for phlogenetic study. SPSS was used by Ba et al. (2002) for phygenetic study of rhizobia. Similarly, GeneSpring 7.3.1 was used by Koch et al. (2010). Other tools and packages that are available for phylogenetic tree construction are Cluster 3.0, ELKI, Octave, Orange, SCaVis, Scikit-learn, Weka, and CrimeStat. There are several evidences of using hierarchical clustering for phylogenetic tree creation in literature but the name of the algorithm has not been authors (Mathur and Tuli 1990; Frédéric Ampe et al. 2003; Korner et al. 2003; Bontemps et al. 2005; Capoen et al. 2007; Brechenmacher et al. 2008; Schuller et al. 2012; Choi and Yun 2016).

10.7 Phylogenetic Tools Used for Rhizobial Research (1990–1999)

Phylogenetic analysis of rhizobia and agrobacteria was performed by Willems and Collins (1993) using 16s RNA gene sequences obtained from EMBL Data Library. Tools used for pairwise sequence analysis and phylogenetic tree construction have been discussed in Table 10.2. Results of phylogenetic analysis suggested that the genera *Bradyrhizobium* and *Azorhizobium* belong to distinct phylogenetic lineages, and there is evidence of intermixing of *Rhizobium* and *Agrobacterium* species in subgroups. Phylogenetic relationships among *Rhizobium* species for nodulating the common bean (*Phaseolus vulgaris* L.) was determined by Berkum et al. in 1996. A direct sequencing of amplified 16s ribosomal DNA genes was performed. Tools used for alignment of sequences, creation, and analysis of phylogenetic trees have been discussed in Table 10.2. As a result, four clusters were formed—cluster 1 with

Table 10.2 The phylogenetic tools used for rhizobial research (1990–1999)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
A. clevelandensis, *A. felis*, *A. tumefaciens*, *B. bacillifomis*, *B. denitrificans*, etc.	16S rRNA gene sequences	EMBL Data Library	Genetics Computer Group Sequence Analysis package V.7.01	Sequence similarities for pairwise alignments	Not mentioned	Willems and Collins (1993)
			DNAPARS and DNABOOT of the Phylogeny Inference Package	Unrooted phylogenetic tree	Parsimony and bootstrap methods insertions/deletions of more than 1 base length	
Rhizobium species nodulating the common Bean (*Phaseolus vulgaris* L.)	16S rRNA	Not mentioned	SEQBOOT, DNADIST, FITCH, and CONSENSE	Alignment and analysis of sequences	Similarity maximum 99.8% to minimum 97.1% with 3 and 41 nucleotide differences	Berkum et al. (1996)
			DRAWTREE and RETREE	Construction and analysis of phylogenetic trees	Jukes–Cantor model and Fitch–Margoliash method	
			Neighbour Program of Felsenstein's Phylip 3.5	Phylogenetic relationships	Neighbour-joining algorithm	
			RETREE	To reroot the constructed tree	Not mentioned	
Not mentioned	SSU rRNA sequences	Not mentioned			Neighbour-joining	Young and Haukka (1996)

Mesorhizobium tianshanense and related rhizobia	rDNA, 16s rRNA	EMBL, GenBank, and DDBJ data libraries	DNADIST of PHYLIP version 3.572	Infer similarities between sequences	Jukes–Cantor coefficient Deletions and insertions of more than one base length	Tan et al. (1997)
			NEIGHBOUR	Create dendrogram	Neighbour-joining method	
			DRAWTREE	Draw unrooted tree	Not mentioned	
Rhizobium etli and other *Rhizobium* spp	16s rRNA, 16s rRNA–23s rRNA intergenic spacer	Not mentioned	SAHN	Generate dendrograms	Not mentioned	Sessitsch et al. (1997)
A. tumefaciens	16S rRNA and 16S-23S rRNA spacer	EMBL	Phylo-Win	For phylogenetic analysis	Neighbour-joining with Kimura and parsimony methods Bootstrapping with 1000 replicates	Khbaya et al. (1998)
			NJplot	Phylogenetic tree	Not mentioned	
			PILEUP	Alignment of 16s and 23s RNA sequences	Not mentioned	
			CLUSTALW	Phylogenetic tree	Not mentioned	
Mesorhizobium huakuii and *Rhizobium galegae*	*nodB*, *nodC*, GSII 16S rRNA	GenBank	CLUSTALW	Multiple alignments	Gonnet distance matrix	Yang et al. (1999)
			TREEVIEW	Visualize phylogenetic tree	Bootstrap values at the branch points and scale bar. 0.01 substitutions per site	
			PAUP, version 3.1.1	Infer phylogenetic tree	Maximum parsimony	
			MEGA	Neighbour-joining analyses	Kimura's two-parameter method nucleotide distances bootstrapping (100 replicates)	

Rhizobium leguminosarum bv. trifolii, *R. leguminosarum bv. viciae*, and *R. leguminosarum bv. phaseoli*. Cluster 2 and cluster 3 which comprises *Rhizobium etli* and *Rhizobium tropici*, and cluster 4 contained a single bean-nodulating strain (Berkum et al. 1996). Genetic and phylogenetic study of four *Rhizobium* genera was performed by Young and Haukka (1996). Phylogenietic tree of rhizobia and some related bacteria was created by the neighbour-joining method from SSU rRNA sequences and subdivided rhizobia into three genera: *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* that lie in distinct branches of subdivision of the *Proteohacteria* that contains many non-rhizobial bacterial species. Results revealed that the common rhizobial ancestor does not contain genes for legume nodules but procured by phylogenetically distinct bacteria in course of evolution. In essence, nitrogen fixation genes are often linked to nodulation genes, but it need not to have the same evolutionary history. Tan and colleagues have studied the phylogenetic relationships of *Mesorhizobium tianshanense* with other related rhizobia (Tan et al. 1997). The details of phylogenetic tools used for the study have been given in Table 10.2. A clear difference was appeared between *M. tianshanense* cluster and *Rhizobium* cluster for SDS-PAGE.

The DNA–DNA relatedness between type strain of *M. Tianshanense* and type or reference strain of *Mesorhizobium loti*, *M. huakuii*, *M. ciceri*, and *M. Mediterraneum* ranged from 4.4 to 43.8%. Phylogenetic analysis based on the 16s rRNA gene sequences showed that *M. tianshunense* was closely related to the *Mesorhizobium* but distinguished from the other four species in this branch. These results further confirmed that these bacteria constitute a distinct rhizobial species (Tan et al. 1997). The characterization of *R. etli* and other *Rhizobium* spp. was performed by Sessitsch et al. (1997) using PCR analysis with repititive primers that nodulate *P. vulgaris* in Australian soil. The plasmid profiles, *nifH* profiles, PCR-RFLP analysis of 16s rRNA gene, and of the 16s rRNA–23s rRNA intergenic spacer and nodulation phenotypes were analysed. Dendograms were generated using SAHN and results suggested that *Phaselous vulgaris* strain found in Austria were derived from rhizobia obtaining in Mesoamerica (Sessitsch et al. 1997). The genetic diversity and phylogeny of 40 rhizobia that nodulating four Acacia species viz. *A. Gummifera*, *A. Raddiana*, *A. Cyanophylla*, and *A. Horrid* from Morocco were analysed by Khbaya et al. (1998) using rRNA and 16S–23S rRNA spacer by PCR with RFLP analysis. Tools used for phylogenetic analysis are discussed in Table 10.2. 16s RNA analysis identified three clusters out of which two belonging to *Sinorhizobium meliloti* and *Sinorhizobium fredii*. The third cluster was *Rhizobium galegae* that is closely related to the *Agrobacterium tumefaciens* species whose phylogenetic position was determined with respect to other rhizobia and agrobacteria using PCR-RFLP with nine restriction enzymes of 23s rRNA genes of 42 rhizobial and agrobacterial strains retrieved from the EMBL database. As a result, 27 and 32 different restriction patterns were found for 16s and 23s RNA which were aligned using PILEUP and a phylogenetic tree was constructed using CLUSTALW. The 16S analysis of *R. galegae* formed a sub-group on the *Agrobacterium* branch, but in the 23s analysis, they are part of the *Rhizobium* branch (Khbaya et al. 1998).

The *nod* gene of the *Mesorhizobium huakuii* and *R. galegae* was studied by a, b-unsaturated *N*-acyl substitutions (Yang et al. 1999). The in silico tools used for this analysis are discussed in Table 10.2. The benchmarking of the evolutionary dynamics of symbiotic and housekeeping loci of the genetic coherence of rhizobial lineages was performed by isolating 47 rhizobial strains from nodules of 13 genera of the temperate herbaceous *Papilionoideae* across several continents. Analysis showed that each locus subdivides strains into genera *Rhizobium*, *Sinorhizobium,* and *Mesorhizobium*. In contrast to the previous study, results indicate a lack of lateral transfer across major chromosomal subdivisions and a significant incongruence of *nod* and GSII phylogenies within rhizobial subdivisions which strongly suggests horizontal transfer of *nod* genes among congenerics (Yang et al. 1999).

10.8 Phylogenetic Tools Used for Rhizobial Research (2000–2010)

A study of nitrogen-fixing nodules of *Ensifer adhaerens* harbouring *R. tropici* symbiotic plasmids was performed (Rogel et al. 2001). The ribosomal fingerprinting was performed digesting PCR products with 16S rRNA gene restriction enzyme *Hinf*I, *Msp*I, *Rsa*I, *Hha*I, *Sau*3A1, and *Dde*I with primers fD1 and rD1 from *E. adhaerens* transconjugants. The details of in silico analysis are given in Table 10.3. Results indicated that *E. adhaerens* is related to *Sinorhizobium* spp. *E. Adhaerens* did not nodulate *P. vulgaris* (bean) or *Leucaena leucocephala*, but with symbiotic plasmids from *R. tropici*, it formed nitrogen-fixing nodules on both hosts. A close relationship among *P. vulgaris* symbionts was revealed on classifying a collection of 83 rhizobial strains based on *nodC* and *nifH* genes in 23 recognized species distributed in the genera *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Bradyrhizobium*, as well as unclassified rhizobia from various host legumes. Irrespective of 16S rRNA-based classification, phylogenetic trees revealed that *nodC* and *nifH* were similar but incongruence in some cases suggested that genetic rearrangements have occurred in course of evolution. This is an indication of lateral genetic transfer across *Rhizobium* and *Sinorhizobium* genera that played a role in diversification and in structuring of population of rhizobia (Rogel et al. 2001).

Velázquez et al. (2005) worked on the coexistence of symbiosis and pathogenicity-determining genes in *Rhizobium rhizogenes* strains that enabled them to induce nodules and tumours or hairy roots in plants. The in silico tools are discussed in Table 10.3. *Rhizobium* sequence analysis of 12 rhizobial species was performed using 16S rRNA and *dnaK* genes (Table 10.3) (Eardly et al. 2005). The discordance between 16S rRNA and *dnaK* phylogenies was tested with the incongruence length difference (ILD) test. As a result, two groups of related species were identified by neighbour-joining and maximum parsimony analysis. One group consisted of *M. loti* and *Mesorhizobium ciceri*, and the other group consisted of *Agrobacterium rhizogenes*, *R. tropici, R. etli,* and *R. leguminosarum*. Although

Table 10.3 The phylogenetic tools used for rhizobial research (2000–2010)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/method/parameter used	Author
Rhizobium tropici	16S rRNA, *nodC*, *nifH*	Not mentioned	FITCH of PHYLIP	Generate phylogenetic tree	Distance matrix	Rogel et al. (2001)
			ClustalW	Sequence alignment	Jukes and Cantor method	
			Bisance software	Amplification of *nodC* and *nifH* fragments	Amplification of up to 930 and 780 bp	
			Phylip (Felsenstein 1989	Generate and infer phylogenetic trees	Neighbour-joining, Kimura's two-parameter method, and maximum likelihood	
			Protdist program of Phylip	Phylogenetic tree of *nodC* and *nifH* proteins	Dayhoff PAM distance matrix	
			SEQBOOT and consense programs of PHYLIP	Create neighbour-joining tree	Bootstrap analysis	
Rhizobium rhizogenes		GenBank	BLAST	Sequence comparison	Not mentioned	Velázquez et al. (2005)
			ClustalW	Sequence alignment	Not mentioned	
			MEGA2	Phylogenetic trees	Neighbour-joining method and bootstrap analysis Kimura's two-parameter method to find distances based on 1000 resamplings	

Rhizobium galegae	16S rRNA and *dnaK* Genes	Not mentioned	ClustalW	Alignment of sequences	A two-step process the IUB DNA weight matrix and (for protein sequences) the PAM 250 protein weight matrix	Eardly et al. (2005)
			CodonAlign 2.0	Nucleotide sequence alignment	DnaK amino acid sequence alignment	
			MEGA version 2.1	Neighbour-joining phylogenetic tree creation	Neighbour-joining algorithm Jukes–Cantor distances	
				Maximum parsimony trees	Heuristic min-mini tree search option search factor of 2	
				Bootstrap analysis	Analysing Bootstrap confidence levels 1000 permutations of the data sets	
Not mentioned	Not mentioned	Not mentioned	WebPHYLIP	Plot trees	Not mentioned	Zhang et al. (2007)
			ClustalW	Pairwise alignment	Default settings	
Brazilian *Rhizobium tropici* strains	16S rRNA	Not mentioned	ClustalX version 1.83	Multiple sequence alignments	Not mentioned	Pinto et al. (2007)
			MEGA version 3.1	Generate phylogenetic trees (16S rRNA phylogeny)	Neighbour-joining algorithm and K2P distance model default parameters, *Azospirillum brasilense* as an outgroup	
					Bootstrap analysis 2000 samplings	

(continued)

Table 10.3 (continued)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/method/parameter used	Author
Thirteen *Rhizobium legumnosarum* bv. viciae	DNA	Not mentioned	BLAST	Alignment of homologous proteins	JGI locus tags Ne0441 to Ne0457	Blažinkov et al. (2007)
			UPGMA	Hierarchical cluster analysis to construct a dendrogram	Not mentioned	
			BLAST	Sequence matching	Not mentioned	
			MultAlin TCoffee SeaView	Multiple alignments and manual editing of sequences	Not mentioned	
			Gblocks	Extract unambiguously aligned sequence blocks	Default parameters	
			ProtTest 1.3	Select the best model of protein evolution	Substitution matrices—WAG, RtREV, and Blosum62	
			PAL2NAL	Convert amino acid alignments to nucleotide alignments	Gblocks	
			PHYML 2.4.4	Maximum likelihood analyses	HKY and GTR models of protein and nucleotide evolution	
				Nonparametric analysis	Bootstrap analysis 100 replicates	
				Approximate likelihood ratio test	Used as branch support measures	
Different 30, 17, 25 species	16S rRNA and protein sequences of NifH, LuxA, and LuxS	Not mentioned	GeneBee, ClustalW, and Phylip	Phylogenetic analysis	Cladograms, phylograms, and unrooted radial trees are generated	Chaphalkar and Salunkhe (2010)

bootstrap support for the placement of the remaining six species varied, *A. tumefaciens*, *A. rubi*, and *A. vitis* were consistently associated in the same sub-cluster. The three other species included were *R. galegae*, *S. meliloti*, and *Brucella ovis*. The placement of *R. galegae* was the least consistent in this study. It was placed flanking the *A. rhizogenes-Rhizobium* cluster in the *dnaK* nucleotide sequence trees. On the other hand, it was placed with the other three *Agrobacterium* species in the 16S rRNA and the DnaK amino acid trees. An effort to explain the inconsistent placement of *R. Galegae* was performed by examining the polymorphic site distribution patterns among the various species. The similarity in localized runs of nucleotide sequence was an evident and suggesting that the *R. galegae* genes are chimeric. These results provide a tenable explanation for the phylogenetic placement of *R. galegae*, and they also illustrate a potential pitfall in the use of partial sequences for species identification (Eardly et al. 2005).

An attempt was performed for monophyletic clustering and characterization of protein families of *M. tuberculosis*, *Rhizobium sp.*, *E. coli*, *H. pylori*, *Synechocystis sp.*, *M. thermoautotrophicum*, *A. aeolicus*, *B. burgdorferi, P. horikoshii, T. pallidum, B. subtilis, M. jannaschii, H. influenzae,* and *A. fulgidus* was made (Zhang et al. 2007) (Table 10.3). A polyphasic characterization of Brazilian *R. tropici* strains effective in fixing N_2 with common bean (*P. vulgaris* L.) was done (Pinto et al. 2007). Phylogenetic analysis was performed using tools indicated in Table 10.3. The results have shown that the trend of a group of monophyletic proteins might be characterized by a normal distribution, while the strength and variability of this trend can be described by the sample mean and variance of the observed correlation coefficients after a suitable transformation. Genotypic characterisation of indigenous *R. leguminosarum* was performed (Blažinkov et al. 2007). Thirteen *R. leguminosarum bv. viciae* strains were isolated from continental part of Croatia and were analysed using two DNA fingerprinting methods, Randomly Amplified Polymorphic DNA (RAPD-PCR) and Repetitive Extragenomic Palindromic-PCR (REP-PCR). The UPGMA algorithm was used to perform hierarchical cluster analysis and to construct a dendrogram. An evolution and functional characterization of the RH50 gene from the ammonia-oxidizing bacterium *Nitrosomonas europaea* was performed. For phylogenetic analysis, various tools are used that are discussed in Table 10.3. Analysis with nonparametric bootstrap analysis and an approximate likelihood ratio test, both methods resulted in similar grouping of strains. Cluster analysis of REP and RAPD-PCR profiles showed significant differences among *R. leguminosarum bv. viciae* isolates. These results suggested the presence of adapted indigenous *R. leguminosarum bv. viciae* strains, probably with higher competitive ability, whose symbiotic properties were evaluated (Blažinkov et al. 2007).

Phylogenetic analysis of nitrogen-fixing and quorum-sensing bacteria was performed (Chaphalkar and Salunkhe 2010). Protein sequences of NifH (nitrogenase reductase), LuxA (Luciferase alpha subunit), and LuxS (Sribosyl homocysteine lyase) from 30, 17, and 25 species of bacteria were aligned, respectively. Phylogenetic analyses on the basis of 16S rRNA was performed using GeneBee, ClustalW, and PHYLIP. Further details are given in Table 10.3. Phylogenetic

trees were constructed in the form of cladograms, phylograms, and unrooted radial trees. According to the results obtained, the most highly evolved group of organisms with respect to their nitrogenase reductase protein is that of *Desulfovibrio vulgaris* and *Chlorobium phaeobacteriodes*. *Bacillus thuringiensis* and *Bacillus subtilis* hold the most highly evolved forms of LuxS protein. The motif pattern analysis between *Bradyrhizobium japonicum* and *R. leguminosarum* NifH protein sequence shows that there may be quorum-sensing mediated gene regulation in host bacterium interaction (Chaphalkar and Salunkhe 2010).

10.9 Phylogenetic Tools Used for Rhizobial Research (2011–2016)

The genetic diversity of rhizobia-nodulating lentil (*Lens culinaris*) in Bangladesh was performed by phylogenetic analysis of housekeeping genes (16S rRNA, *recA*, *atpD*, and *glnII*) and nodulation genes (*nodC*, *nodD*, and *nodA*) of 36 bacterial isolates from 25 localities across the country (Rashid et al. 2012). BioEdit, Mega, and MrBayes were used for alignment and tree construction and analysis (Table 10.4). Results indicated that most of the isolates (30 out of 36) were related to *R. etli* and *R. leguminosarum*. Only 30 isolates were able to re-nodulate lentil under laboratory conditions. The protein-coding housekeeping genes of the lentil-nodulating isolates showed 89.1–94.8% genetic similarity to the corresponding genes of *R. etli* and *R. leguminosarum*. The same analyses showed that they split into three distinct phylogenetic clades (Rashid et al. 2012).

A characterization of rhizobia-nodulating *Galega officinalis* and *Hedysarum coronarium* was performed (Liu et al. 2012). The study indicated that these species of New Zealand form effective nodules with *R. galegae* and *R. Sullae* only. The sequence analysis of 16S rRNA and housekeeping genes and plant nodulation tests were carried out. Only *R. galegae* strains were isolated from *G. officinalis* and selected strains induced effective nodules when re-inoculated onto the host plant. *Agrobacterium vitis*, *R. galegae*, and *R. sullae* strains were isolated from nodules of *H. coronarium*, but only *R. sullae* induced effective nodules on this plant. For phylogenetic analyses, DNA sequences were aligned and Maximum Likelihood (ML) trees were constructed with 1000 bootstrap replications using MEGA5 software (Table 10.4). Model test was performed and the best model was selected for each gene. The models of evolution used for 16S rRNA, *atpD*, and *recA* were T92+G+I, T92+I, and T92+G, respectively. Results from this study concur with previous reports on their high degree of specificity in relation to their rhizobial symbionts. *Mesorhizobium* spp. known to nodulate New Zealand native legumes were not found in the nodules of *G. officinalis* and *H. coronarium*. However, further work, which included cross-nodulation tests with native rhizobia and sampling of both legumes at various sites, would confirm the specificity of these legumes in New Zealand (Liu et al. 2012). A discovery of a new beta-proteobacterial

Table 10.4 The phylogenetic tools used for rhizobial research (2011–2016)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
Rhizobia-nodulating lentil (*Lens culinaris*) in Bangladesh	Housekeeping genes (16S rRNA, *recA*, *atpD*, and *glnII*) and nodulation genes (*nodC*, *nodD*, and *nodA*)	36 bacterial isolates from 25 localities across the country	BioEdit	Multiple alignment	Not mentioned	Rashid et al. (2012)
			MEGA version 5	*p*-distance	Not mentioned	
				Phylogenetic tree creation and analysis	Neighbour-joining (NJ) algorithm and maximum likelihood Kimura two-parameter model (K2P)	
				Bootstrap Analysis	Bootstrap support with 1000 replicates	
				Tree construction	All trees rooted with Brady*rhizobium* as outgroup Trees sample = every 500 generations burn in = first 4000 samples(discarded)	
			MrBayes version 3.1.2	Phylogenetic inference	Bayesian Inference (BI), runs = two independent, generations = 8,000,000, Markov chains = 4	
Rhizobium galegae and *R. Sullae*	16S rRNA and housekeeping genes and DNA		MEGA5	DNA sequences were aligned and maximum likelihood (ML) trees	Maximum likelihood 1000 bootstrap replications	Liu et al. (2012)

(continued)

Table 10.4 (continued)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
New Beta-*Rhizobium*-nodulating *Parapiptadenia rigida* (Benth.)	*nifH* and 16S rRNA genes	47 isolates	Greengenes program using the NAST alignment tool	Nucleotide alignments of 16S rRNA	Manually edited	Taulé et al. (2012)
			CLUSTALW version 1.8	Nucleotide alignments of the *nifH*, *nodA*, and *nod* sequences		
			MEGA4	Phylogenetic trees	Neighbour-joining algorithm Kimura two-parameter substitution model, 1000 bootstrap replications for bootstrap consensus tree	
			Psi-BLAST	T3SS core protein sequences	With the *P. syringae pv phaseolicola* 1448a T3SS-2 gene cluster coding frames	
Rhizobium pongamiae sp. from root nodules of Pongamia pinnata	16S rRNA, *rec*A, and *atp*D genes	GenBank	BLASTN	Compare sequences	Not mentioned	Kesari et al. (2013)
			ClustalW2	Multiple sequence alignment	Not mentioned	
			MEGA 4.0	Phylogenetic trees	Bootstrap analysis 1000 resamplings	
					Neighbour-joining method Kimura-2 model	

Rhizobia from *Arachis hypogaea* L.	Not mentioned	Not mentioned	NTSYS pc version 2.01	Similarity matrix	Not mentioned	Lyra et al. (2013)
			UPGMA	Cluster analysis	Genetic distances Simple matching coefficient (SM)	
				Dendrograms	SAHN method	
			MUSCLE	Protein alignments	Not mentioned	
			PHYLIP	To construct phylogeny	Not mentioned	
Rhizobium leguminosarum bv. Trifolii	Not mentioned	Not mentioned	MEGA, version 5.05	Phylogenetic analyses	Maximum likelihood method General Time Reversible model	Reeve et al. (2013)
					Bootstrap analysis 500 replicates	
Rhizobium grahamii	CCGE502 genome	Not mentioned	JSpecies	Sequence comparison	Not mentioned	Althabegoiti et al. (2014)
			CLUSTALX version 1.83	Multiple sequence alignments	Not mentioned	
			BioEdit	Multiple sequence alignments	Not mentioned	
			ProtTest 2.4	Best fit models of evolution for each gene	Akaike information criterion	
			PhyML 3	Maximum likelihood phylogenies	Subtree pruning and regrafting moves	
				Shimodaira–Hasegawa-like approximate likelihood ratio test	Tree nodes	

(continued)

Table 10.4 (continued)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
Rhizobia isolated from nodules of *Centrolobium paraense*	Not mentioned	Not mentioned	Mega 5.05	Phylogenetic analysis	Neighbour-joining method	Baraúna et al. (2014)
Rhizobium phaseoli and one S. americanum	*rpoB* sequences	Not mentioned	ClustalW	Sequence alignment	Not mentioned	Mora et al. (2014)
			PHYLIP	Infer phylogeny	Not mentioned	
			NJplot	Generate trees	Not mentioned	
Narrow-host-range bacteriophages that infect *Rhizobium etli*		Nonredundant (nr) GenBank and Phage Orthologous Group (POG)-10 database	BLASTX	Sequence alignment	MCL algorithm against the terminases of *R. etli* phages	Santamaría et al. (2014)
			Phred/Phrap/ Consed software package	DNA sequencing and assemble reads	Not mentioned	
			Glimmer (version 3.0)	ORFs prediction	Not mentioned	
			ARTEMIS	Annotate genome sequences	With the help of BlastX	
			InterProScan	Searches for putative conserved domains	Against (POG)-10 database	
			MAUVE	Additional comparisons	Conserved blocks among the phage genomes	

Narrow-host-range bacteriophages that infect *Rhizobium* etli	Twenty rhizobial strains isolated from the root nodules of soybean (*Glycine max* L.) from Egypt	16S rDNA, nifH, nodA	DNASTAR	Sequence assembly	Not mentioned	Youseif et al. (2014)
			BLASTN	Sequence similarity searches	Not mentioned	
			ClustalW version 1.8	Align sequences	Not mentioned	
			PHYLIP	Phylogenetic analysis	Neighbour-joining (NJ) method 1000 bootstrap replication	
Native rhizobia-nodulating *Phaseolus lunatus*	DNA, 16S rRNA	Fourteen isolates of rhizobia	BLAST	Preliminary species assignment		Araujo et al. (2015)
			DNAMan version 4.0	Pairwise comparisons	Optimal alignment option; k-tuple = 2, gap penalty = 7, gap open = 10, and gap extension = 5	
			ClustalW	Alignment of nucleotide sequences	Not mentioned	
			MEGA v. 6.0.	Generate and infer phylogenetic tree	Neighbour-joining (NJ) algorithm and maximum likelihood (ML) methods Kimura two-parameter distance correction model; evaluated bootstrap support for each node using 1000 replicates	

(continued)

Table 10.4 (continued)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
Rhizobium from nodulating beans grown in Mediterranean climate soils of Chile	DNA	GenBank	UPGMA	Cluster analysis and dendrograms creation	Not mentioned	Baginsky et al. (2015)
			CLUSTALX	Sequence alignment	Eight sequences from Chile and 24 from Genbank	
			Mesquite 2.75	Visual inspection of sequences	Not mentioned	
			MEGA5.2	Create and infer phylogenetic trees	Neighbour-joining and maximum likelihood	
			Model test	Select evolutionary model	Clades with 1000 bootstrap replicates	
			Tree View	Visualize phylogenetic trees		
Rhizobia isolated from 3 Tunisian wild legume species of the genus sulla	16S rRNA gene and ITS region sequences		BLASTN	Construct two data sets(from *Agrobacterium* and *Rhizobium*)	First data set—20 16S rRNA sequences second data set contained 21 ITS seq	Chriki-Adeeb and Chriki (2015)
			ClustalX version 2.0.10	Alignment of data sets	1513 and 1636 nucleotide positions	

			MrBayes program V3.2	Create baysian phylogenetic tree	Best-fit model of nucleotide substitution HKY substitution model	
					Bayesian MCMC method; generations = 1 million matrix = HKYmodel parameters = (gamma shape and proportion invariant) sample trees = every 500 generations (default value)	
Rhizobium leguminosarum	Not mentioned	Not mentioned	SPLITSTREE v. 4.11	Create neighbour-nets	Uncorrected *p*-distances function	Kumar et al. (2015)
				Pairwise homoplasy index test	Applied to each of the 100 genes with 5% significance level	
			FASTTREE	Maximum likelihood analyses	Gamma–gtr option	
				Create ML tree	100-gene alignment FASTTREE with 100 bootstrap replicates	
			PHYML	ML phylogeny	Best-fit model of nucleotide substitution	
			MODELTEST in TOPALI v. 2	Find best-fit model	Not mentioned	
			CONSEL	Congruence test	*p*, 0.05: incongruent	
			R package PHYLCON	Infer phylogenetic trees	Heatmaps to display *p*-values of SH test	

(continued)

Table 10.4 (continued)

Species	Sequence	Database used	Tools/Program used	Purpose	Algorithm/coefficient/ method/parameter used	Author
Rhizobium sullae	16S rRNA, recA, nodD, and nifH genes	Not mentioned	Muscle	Multiple nucleotide sequence alignments	Not mentioned	Aliliche et al. (2016)
			MEGA version 6	Phylogenetic analysis	Maximum likelihood methods Bootstrap analyses using 1000 replicates branching point = C70% bootstrap value	
Rhizobium vitis	Transcriptional profiles of *Rhizobium vitis*	Not mentioned	Hierarchical clustering	Cluster analysis in tree creation	Euclidean distances normalized significant genes	Choi and Yun (2016)
			Avadis Prophetic Ver. 3.3	Analyse patterns of expressed changes	Not mentioned	

Rhizobium strains was performed in (Taulé et al. 2012), which was able to efficiently nodulate *Parapiptadenia rigida (Benth.) Brenan*.

A collection of Angico-nodulating isolates was obtained and 47 isolates were selected for genetic studies. According to entero-bacterial repetitive intergenic consensus PCR patterns and RFLP analysis of their *nifH* and 16S rRNA genes, the isolates could be grouped into seven genotypes, including the genera *Burkholderia*, *Cupriavidus*, and *Rhizobium*, among which the *Burkholderia* genotypes were the predominant group. Details of the tools used for this study was given in Table 10.4. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the amino acid sequences analysed. Branches corresponding to partitions reproduced in <50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) has been shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated in pairwise sequence comparisons. Phylogenetic studies of *nifH*, *nodA*, and *nodC* sequences from the *Burkholderia* and the *Cupriavidus* isolates indicated a close relationship of these genes with those from beta-proteobacterial rhizobia (beta-rhizobia) rather than from alpha-proteobacterial rhizobia (alpha-rhizobia). In addition, nodulation assays with representative isolates showed that while the *Cupriavidus* isolates were able to effectively nodulate *Mimosa pudica*, the Burkholderia isolates produced white and ineffective nodules on this host (Taulé et al. 2012). *Rhizobium pongamiae sp. from root nodules of Pongamia pinnata* was studied in (Kesari et al. 2013). Phylogenetic analysis of sequences of 16S rRNA, *rec*A, and *atp*D genes was performed using tools discussed in Table 10.4. Phenotypic and molecular study of rhizobia isolated from nodules of peanut (*Arachis hypogaea* L.) grown in Brazilian Spodosols (Pernambuco State) was performed (Lyra et al. 2013). A total of 22 bacterial strains were isolated from nodules of seven peanut varieties. Refer Table 10.4 for details. The genome sequence of the clover-nodulating *Rhizobium leguminosarum bv. trifolii* strain TA1 was analysed (Table 10.4) (Reeve et al. 2013). A little information about the phylogeny of the isolates was found by the analysis of the phenotypic characteristics-colony morphology and IAR. A great diversity of these rhizobia and the presence of new species were revealed by using compilation of phenotypic and molecular characteristics.

The genome sequence and transfer properties of *Rhizobium grahamii* was studied (Althabegoiti et al. 2014). The *Genome* sequence was obtained from *R. grahamii* CCGE502 type strain isolated from *Dalea leporina* in Mexico. It comprises one chromosome and two extrachromosomal replicons (ERs), pRgrCCGE502a and pRgrCCGE502b, and a plasmid integrated in the CCGE502 chromosome. Several analysis tools were used for phylogenetic study. Details of these tools are presented in Table 10.4. The analysis showed variable degrees of nucleotide identity and gene content conservation in *R. grahamii* CCGE502

replicons as compared to *R. mesoamericanum* genomes. The extrachromosomal replicons from *R. grahamii* were similar to those found in other related *Rhizobium* species. A limited similarity was observed in *R. grahamii* CCGE502 symbiotic plasmid and megaplasmid in distant *Rhizobium* species. The set of conserved genes in *R. grahamii* are highly expressed in *R. phaseoli* on plant roots. This was an indication of its role in root colonization. The diversity and nitrogen fixation efficiency of rhizobia isolated from nodules of *Centrolobium paraense* was studied (Baraúna et al. 2014). Soil samples were collected from four sites of the Roraima Cerrado, Brazil and used to cultivate *C. paraense* in order to obtain nodules. The results revealed that *C. paraense* is able to nodulate with different *Rhizobium* species and *Bradyrhizobium* isolates had the highest symbiotic efficiency on *C. Paraense* and showed a contribution similar to the nitrogen treatment, some of which have not yet been described. The nitrogen-fixing rhizobial strains were isolated from non-inoculated bean plants. Total nine isolates were obtained which belong to the *Rhizobium* and *Sinorhizobium* groups. The strains showed several large plasmids, except for a *Sinorhizobium americanum* isolate (Table 10.4) (Mora et al. 2014). Fourteen narrow-host-range bacteriophages that infect *R. etli* were isolated from rhizosphere soil of bean plants from agricultural lands in Mexico using an enrichment method (Santamaría et al. 2014). The complete genome of nine phages of size varied from 43 to 115 kb was obtained. Four phages were resistant to several restriction enzymes. A large proportion of open reading frames of these phage genomes (65–70%) consisted of hypothetical and orphan genes. Refer Table 10.4 for details of in silico tools used in this study. Authors have classified these phages into four genomic types on the basis of their genomic similarity, gene content, and host range and proposed that these bacteriophages correspond to novel species (Santamaría et al. 2014).

Twenty rhizobial strains isolated from the root nodules of soybean (*Glycine max* L.) were collected from diverse agro-climatic and soil conditions in Egypt (Youseif et al. 2014). The strains were characterized using a polyphasic approach, including nodulation pattern, phenotypic characterization, 16S rDNA sequencing, *nifH* and *nodA* symbiotic genes sequencing, and REP-PCR fingerprinting. Please refer Table 10.4 for details. The complete sequencing of 16S rRNA demonstrated that native soybean-nodulating rhizobia are phylogenetically related to *Bradyrhizobium*, *Ensifer*, and *Rhizobium* (syn. *Agrobacterium*) genera. The study of tolerance ability to environmental stresses revealed that local strains survived in a wide pH ranges (pH 5–11) and a few of them tolerated high acidic conditions (pH 4). *Agrobacterium* strains were identified as the highest salt tolerant and were survived under 6% NaCl; however *Ensifer* strains were the uppermost heat tolerant and can grow at 42°C. The DNA and the 16S rRNA gene of 14 isolates of rhizobia-nodulating *Phaseolus lunatus* from Brazil were extracted and sequenced using primers fD1 and rD1 (Araujo et al. 2015). Phylogenetic study was performed using tools discussed in Table 10.4. More than 50% of strains studied were positioned in the *Bradyrhizobium* clade and one strain was positioned in the *R. etli*/*Rhizobium phaseoli* clade. Two strains were grouped within the *R. tropici* group and three strains, ISOL16, ISOL21, and ISOL27 represent new lineages. This

is a clear indication of that there is a high species diversity of rhizobia-nodulating *P. lunatus* in Northeast Brazil, including potential new species. To study the genetic diversity of *Rhizobium* from nodulating beans grown in a Mediterranean climate soils of Chile, the genetic similarity among the PCR-RFLP patterns was performed (Baginsky et al. 2015). The phylogenetic analysis tools used in this study have been presented in Table 10.4. The bayesian phylogenetic analysis of rhizobia of the genus *Sulla* was performed on three Tunisian wild legume species (Chriki-Adeeb and Chriki 2015). The phylogenetic relatedness and substitution rates of 16S rRNA gene and ITS region sequences were analysed by using a relaxed-clock program (Multidivtime) (Table 10.4). The results indicate that Bayesian inferred trees were congruent and showed a clear split between *Agrobacterium* and *Rhizobium* species. The ITS region evolutionary rate was 15-fold higher than the 16S rRNA gene rate, suggesting that the ITS region represented an appropriate molecular marker for inferring phylogenies and divergence times in bacteria. Phylogeny of genospecies of *R. leguminosarum* that are not ecologically coherent was studied by (Kumar et al. 2015). Phylogenetic trees were constructed using either neighbour-net or maximum likelihood (ML) methods. A molecular phylogenetic analysis of *Rhizobium sullae* isolated from Algerian *Hedysarum flexuosum* was performed by (Aliliche et al. 2016) using 16S rRNA, *recA*, *nodD*, and *nifH* genes (Table 10.4). Choi and Yun have analysed transcriptional profiles of *Rhizobium vitis*. Complete linkage hierarchical clustering based on the Euclidean distances of samples was performed using the normalized significant genes. The patterns of expressed changes were analysed for groups using the Avadis Prophetic Ver. 3.3 software (Choi and Yun 2016).

10.10 Conclusion

A number of hierarchical clustering-based algorithms and in silico techniques have been used by researchers for phylogenetic analysis of rhizobia. These popular tools include Blast, Blastn, and BioEdit for pairwise sequence alignment; Muscle, TCofee, ClustalW, and ClustalX for multiple sequence alignment; Phylip tools for phylogenetic inference such as Drawgram to plot rooted tree, DrawTree to draw unrooted tree, consensus to compute consensus tree; MrBayes for Bayesian inference of phylogeny of *Rhizobium*; Mega—a complete package for sequence alignment and phylogenetic inference and UPGMA—a hierarchical algorithm for creating evolutionary tree. We hope the information content from this chapter will help emerging researchers to perform further empirical study to understand rhizobial phylogeny in more details.

References

Abdel-Aziz RA, Al-Barakah FN, Al-Asmary HM (2008) Genetic identification and symbiotic efficiency of *Sinorhizobium meliloti* indigenous to Saudi Arabian soils. Afr J Biotechnol 7: 2803–2809

Aliliche et al (2016) Molecular phylogenetic analysis of *Rhizobium sullae* isolated from Algerian *Hedysarum flexuosum*. Antonie van Leeuwenhoek 109:897–906

Althabegoiti MJ, Ormeño-Orrillo E, Lozano L, Tejerizo GT, Rogel MA, Mora J, Martínez-Romero E (2014) Characterization of *Rhizobium grahamii* extrachromosomal replicons and their transfer among rhizobia. BMC Microbiol 14:6

Araujo ASF, Lopes ACA, Gomes RLF, Beserra Junior REA, Antunes JEL, Lyra MCCP, Figueiredo MVB (2015) Diversity of native rhizobia-nodulating *Phaseolus lunatus* in Brazil. Legume Res 38(5):653–657

Ba S, Willems A, Lajudie PD, Roche P, Jeder H, Quatrini P, Neyra M, Ferro M, Promé JC, Gillis M, Boivin-Masson C, Lorquin J (2002) Symbiotic and taxonomic diversity of *Rhizobia* isolated from *Acacia tortilis* subsp. *raddiana* in Africa. Syst Appl Microbiol 25:130–145

Baginsky et al (2015) Genetic diversity of *Rhizobium* from nodulating beans grown in a variety of Mediterranean climate soils of Chile. Arch Microbiol 197:419–429

Baraúna AC, Silva K, Pereira GMD, Kaminski PE, Perin L, Zilli JE (2014) Diversity and nitrogen fixation efficiency of rhizobia isolated from nodules of *Centrolobium paraense*. Pesq Agropec Bras 49:296–305

Berkum PV, Beyene D, Eardly BD (1996) Phylogenetic Relationships among *Rhizobium* species nodulating the common bean (*Phaseolus vulgaris* L.). Internationjaolu rnaolf systematibca cteriologjyan 46:240–244

Blažinkov M, Sikora S, Uher D, Maćešić D, Redžepović S (2007) Genotypic characterisation of indigenous *Rhizobium leguminosarum* bv. *viciae* field population in Croatia. Agric Conspec Sci 72:153–158

Bontemps C, Golfier G, Gris-Liebe C, Carrere S, Talini L, Boivin-Masson C (2005) Microarray-based detection and typing of the *Rhizobium* nodulation gene nodC: potential of DNA arrays to diagnose biological functions of interest. Appl Environ Microbiol 71:8042–8048

Brechenmacher et al (2008) Transcription profiling of *Soybean* nodulation by *Bradyrhizobium japonicum*. Mol Plant Microbe Interact 21:631–645

Capoen W, Den Herder J, Rombauts S, Gussem JD, Keyser AD, Holsters M, Goormachtig S (2007) Comparative transcriptome analysis reveals common and specific tags for root hair and crack-entry invasion in *Sesbania rostrata*. Plant Physiol 144:1878–1889

Chaphalkar A, Salunkhe N (2010) Phylogenetic analysis of nitrogen-fixing and quorum sensing bacteria. Int J Bioinf Res 2:17–32

Choi YJ, Yun HK (2016) Transcriptional profiles of *Rhizobium vi*tis-inoculated and salicylic acid-treated 'Tamnara' grapevines based on microarray analysis. J Plant Biotechnol 43:37–48

Chriki-Adeeb R, Chriki A (2015) Bayesian phylogenetic analysis of rhizobia isolated from root-nodules of three Tunisian wild legume species of the genus Sulla. J Phylogen Evol Biol 3:149

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. Atlas Protein Seq Struct 5:345–352

Dourado AC, Alves PIL, Tenreiro T, Ferreira EM, Tenreiro R, Fareleira P, Crespo MTB (2009) Identification of *Sinorhizobium* (*Ensifer*) medicae based on a specific genomic sequence unveiled by M13-PCR fingerprinting. Int Microbiol 12:215–225

Eardly BD, Nour SM, Berkum PV, Selander RK (2005) Rhizobial 16S rRNA and dnaK genes: mosaicism and the uncertain phylogenetic placement of *Rhizobium galegae*. Appl Environ Microbiol 71:1328–1335

Faisal T, Farooq J, Vessey K (2009) Genetic diversity of *Bradyrhizobium japonicum* within soybean growing regions of the north-eastern Great Plains of North America as determined by REP-PCR and ERIC-PCR profiling. Symbiosis 48:131–142
Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376
Felsenstein J (1989) PHYLIP – phylogeny inference package (version 3.2). Cladistics 5:164–166
Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284
Frédéric Ampe et al (2003) Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. Genome Biol 4:R15
Hasegawa M, Kishino H, Yano T (1985) Dating of human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22(2):160–174
Hassen et al (2014) Nodulation study and characterization of Rhizobial microsymbionts of forage and pasture legumes in South Africa. World J Agri Res 2(3):93–100
Jia et al (2015) Identification and classification of Rhizobia by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Proteomics Bioinf 8(6):98–107
Jukes TH, Cantor CR (1969) Evolution of protein molecules. Academic Press, New York, pp 21–132
Jurelevicius et al (2010) Polyphasic analysis of the bacterial community in the rhizosphere and roots of *Cyperus rotundus* L. grown in a petroleum-contaminated soil. J Microbiol Biotechnol 20(5):862–870
Kesari et al (2013) *Rhizobium pongamiae sp. nov*. from root nodules of *Pongamia pinnata*. Hindawi Publishing Corporation. BioMed Res Int 2013:65198
Khbaya et al (1998) Genetic diversity and phylogeny of Rhizobia that nodulate *Acacia spp*. in morocco assessed by analysis of rRNA genes. Appl Environ Microbiol 64:4912–4917
Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120
Knief et al (2011) Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. ISME J 6(7):1378–1390
Koch M et al (2010) Rhizobial adaptation to hosts, a new facet in the legume root-nodule symbiosis. MPMI 23:784–790
Korner H et al (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol Rev 792:1–34
Kumar et al (2015) Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. Open Biol 5(1):140133
Liu et al (2012) Characterisation of rhizobia nodulating *Galega officinalis* (goat's rue) and *Hedysarum coronarium* (sulla). N Z Plant Prot 65:192–196
Lyra et al (2013) Phenotypic and molecular characteristics of rhizobia isolated from nodules of peanut (*Arachis hypogaea* L.) grown in Brazilian Spodosols. Afr J Biotechnol 12:2147–2156
Mathur M, Tuli R (1990) Cluster analysis of genes for nitrogen fixation from diazatrophs. J Genet 69:67–78
McGinn et al (2016) *Trifolium* species associate with a similar richness of soil-borne mutualists in their introduced and native ranges. J Biogeogr. doi:10.1111/jbi.12690
Mora et al (2014) Nitrogen-fixing rhizobial strains isolated from common bean seeds: phylogeny, physiology, and genome analysis. Appl Environ Microbiol 80:5644–5654
Murtagh F (1984) Complexities of hierarchical clustering algorithms: the state of the art. Comput Stat Q 1:101–113
Pinto et al (2007) Polyphasic characterization of Brazilian *Rhizobium tropici* strains effective in fixing N2 with common bean (*Phaseolus vulgaris* L.). Soil Biol Biochem 39:1851–1864
Rashid et al (2012) Genetic diversity of rhizobia nodulating lentil (*Lens culinaris*) in Bangladesh. Syst Appl Microbiol 35:98–109
Reeve et al (2013) Genome sequence of the clover-nodulating *Rhizobium leguminosarum bv. trifolii* strain TA1. Stand Genomic Sci 9:243–253

Rogel et al (2001) Nitrogen-fixing nodules with *Ensifer adhaerens* harboring *Rhizobium tropici* symbiotic plasmids. Appl Environ Microbiol 67:3264–3268
Roy SS, Dasgupta R, Bagchi A (2014) A review on phylogenetic analysis: a journey through modern era. Comput Mol Biosci 4:39–45
Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol 10:1073–1095
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Santamaría et al (2014) Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. Appl Environ Microbiol 80:446–454
Schuller et al (2012) Computer-based annotation of putative AraC/XylS-family transcription factors of known structure but unknown function. J Biomed Biotechnol 2012:103132
Sessitsch et al (1997) Characterization of *Rhizobium etli* and other *Rhizobium* spp. that nodulate *Phaseolus vulgaris L.* in an Australian soil. Mol Ecol 6:601–608
Sober E (1983) Parsimony in systematics: philosophical issues. Annu Rev Ecol Syst 14:335–357
Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol Biol Evol 9:678–687
Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526
Tan Z-Y et al (1997) Phylogenetic and genetic relationships of *Mesorhizobium tianshanense* and related Rhizobia. Int J Syst Bacteriol 47:874–879
Taulé et al (2012) New betaproteobacterial *Rhizobium* strains able to efficiently nodulate *Parapiptadenia rigida* (Benth.) Brenan. Appl Environ Microbiol 78(6):1692–1700
Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci Am Math Soc 17:57–86
Tian CF, Zhou YJ, Zhang YM et al (2012) Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc Natl Acad Sci USA 109:8629–8634
Velázquez et al (2005) The coexistence of symbiosis and pathogenicity-determining genes in *Rhizobium rhizogenes* strains enables them to induce nodules and tumors or hairy roots in plants. MPMI 18:1325–1332
Vercruysse et al (2011) Stress response regulators identified through genome-wide transcriptome analysis of the (p) ppGpp-dependent response in *Rhizobium etli*. Genome Biol 12:R17
Willems, Collins (1993) Phylogenetic analysis of *Rhizobia* and *Agrobacteria* based on 16s rRNA gene sequences. Internatiojnoaulr naolf systematbiacc teriologya 43:305–313
Yang G-P et al (1999) Structure of the *Mesorhizobium huakuii* and *Rhizobium galegae* Nod factors: a cluster of phylogenetically related legumes are nodulated by rhizobia producing Nod factors with a,b-unsaturated N-acyl substitutions. Mol Microbiol 34:227–237
Young JPW, Haukka KE (1996) Diversity and phylogeny of rhizobia. New Phytol 8:133
Youseif et al (2014) Phenotypic characteristics and genetic diversity of rhizobia nodulating soybean in Egyptian soils. Eur J Soil Biol 60:34–43
Zhang J et al (2007) Monophyletic clustering and characterization of protein families. J Integr Bioinf 4:67