Pan-genomic Analysis of Bradyrhizobium japonicum

Ruihan Ma[a]
Northeast Agriculture University, Heilongjiang, 15000, China

Keywords: Bradyrhizobium, Biosphere, Genus.

Abstract: The rhizobium-legume symbiosis is a major source of fixed nitrogen (ammonia) in the biosphere—the potential for this process to increase agricultural yield while reducing the reliance on nitrogen-based fertilizers. Bradyrhizobium is an ancient type of soybean nitrogen-fixing symbiotic bacteria. The Bradyrhizobium japonicum under the bacterial genus classification is widely used as a promotional species in actual agricultural production. However, the use of rhizobium also faces some practical problems, like Its nitrogen fixation capacity is unstable. Despite much current research on Bradyrhizobium japonicum, it often focuses on the molecular mechanism of a certain gene or protein, failing to study the environmental adaptability of such bacteria, and failing to study the function and characteristics of bacteria from a genomic perspective. The paper focuses on Pan-genome to study Bradyrhizobium japonicum. Through the obtained genome information tested for line integrity using CheckM and BUSCO, analysis of the Meverage nucleotide consistency (similarity) ANI, and analysis of the secondary metabolite, the genomic dynamics of soybean bradyrhizobium is initially revealed and provides research clues for the analysis of its bacterial functional evolution mechanism and environmental adaptability.

1 INTRODUCTION

The mutually beneficial symbiosis of rhizobium and legumes provides plants with rich nitrogen while actively affecting the soil nitrogen circulation. As a classic model of bionitrogen fixation, with rhizobium jointly regulated by plant roots and rhizobium, its nitrogen fixation occupies more than 60% of the total bionitrogen fixation (Herridge, 2008) and greatly alleviates the nitrogen demand in agriculture. It is worth mentioning that rhizobium

Compound agents (Bradyrhizobium japonicum and Bacillus subtilis) developed by Indigo (Indigo Ag, Inc., Charlestown, USA) can increase crop yield (> 3%) and water absorption efficiency under drought stress (> 75%) with less nitrogen fertilizer application. The success of this model has also made the research and promotion of new bacterial agents or fertilizers the future development direction of agriculture. Rhizobium is a special plant tissue formed by rhizobium and legumes, a process that involves many chemicals, such as plant-secreted flavonoids, isoflavones, and terpenoids (Stoksta, 2016). Some certain root secretions (flavonoids-like, isoflavones) can bind to the NodD protein secretion of rhizobium, causing nod gene expression to produce tuberoma factor, which then acts with the plant root cells to activate tuberoma-related gene expression to form rhizobium (BROGHAMMER, 2012). The genus rhizoma was established in 1889 by B. Frank, containing three species: pea rhizobium, alfalfa, and passion. Bradyrhizobium D.C. Jordan differentiated from rhizobium in 1982. Bradyrhizobium is an ancient soybean nitrogen-fixing symbiotic bacteria widely distributed in different habitats and symbiotic with different legume-specific hosts. Therefore it is highly cosmopolitan (SPRENT, 2017). Bradyrhizobium diazoefficiens USDA110 can form rhizoma with soybean and has excellent symbiotic nitrogen fixation properties. The Bradyrhizobium japonicum under the bacterial genus classification is widely used as a promotional species in agricultural production. Interestingly, as taxonomy developed, the well-known Bradyrhizobium japonicum USDA 110 was eventually divided into Bradyrhizobium diazoefficiens and named Bradyrhizobium diazoefficiens USDA 110. However, the use of rhizobium also faces some practical problems. The

[a] https://orcid.org/0000-0001-5028-5620

Ma, R.
Pan-genomic Analysis of Bradyrhizobium japonicum.
DOI: 10.5220/0011376000003443
In *Proceedings of the 4th International Conference on Biomedical Engineering and Bioinformatics (ICBEB 2022)*, pages 1031-1040
ISBN: 978-989-758-595-1

specific identification of host-rhizomomyces causes rhizobium to be unable to colonize all legume crops. Their nitrogen fixation capacity is good or bad, not as predictable as chemical fertilizers. On the one hand, this is related to rhizobium's genomes and involves the interaction between rhizobium and plants and indigenous microorganisms. Traditionally, most researchers focus on the individual gene or protein to enhance the nitrogen fixation capacity of the rhizobium and rare to focus on genomes, but the effect was not ideal. This paper adopts a new research method, which is to use Pan-genome to study Bradyrhizobium. The Pan-genome is defined as the entire non-redundant gene bank that constitutes the genome, including the core genome, a set of genes (almost) present in all genomes; accessory genome, present in more than two genomes; unique genome, found only in a bacterial genome. In this paper, Pan-genome analysis focuses on Bradyrhizobium, first with identifying core genes, accessory genes, unique genes, and this-based gene function analysis; secondly, the openness and closure of the genome are also important concepts. This paper analyzes the genome sequence of all 21 Bradyrhizobium japonicum and one Bradyrhizobium diazoefficiens USDA 110(from the NCBI due to historical naming).

According to the Heap rule, the closed Pan genome contains all possible genes, and even increasing the scale of genome sequencing, only a small amount of genes are added to the Pan genome.For an open Pan genome, the sequencing of the new genome will increase a lot of undiscovered genes, where its Pan genome is open (L R, V M, P-E F, 2015).Usually, bacterial biogenomes with multiple hosts or frequent habitat changes are more open because their gene islands are more varied; Once specialized intracellular, pathogens are not in constant contact with other bacteria and lose large amounts of genes in evolution to fully adapt to the host. Thus, their genomes are very compact and more closed (BARSY, 2016). The differentiation of the genus rhizobium dates back 200 million years ago (L M, A M, B D, 2001). However, beumes occurred 60 million years ago (MATT, 2005). This and its mismatch suggest that rhizobium's symbiotic nitrogen fixation capacity occurs in modern times, possibly caused by the horizontal transfer of genes. However, whether current genomes are open and how to open rhizobium is still worth studying. The study of the general genome of the rhizome helps to deepen the understanding of the soybean-rhizome symbiosis, provides a theoretical basis for the agricultural production and application of rhizome, and is of great significance to the study of the global nitrogen ecological cycle.

2 METHODS

2.1 Data Source

The genomic data for 21 Bradyrhizobium japonicum and 1 Bradyrhizobium diazoefficiens analynome database and the zed in this paper are from NCBI Gedata format of fasta, is shown in Table 1.In genome assembly, it is assembled from contig into scaffold,contig represents the consensus sequences found from short reads obtained from large-scale sequencing.The first step in assembly is the assembly of contig. from a pair-end library.Further based on mate-pair libraries of different lengths, the originally isolated contig are connected in order, this step yielding scaffolds. Finally, scaffold merged adjusted based on genetic or optical maps to form a chromosome.

Table 1: The strain information used in the study analysis.

species	Bacterial strain	BioSample number	Assembly level	Genome Size (Mb)	GC%	Scaffolds number	Host information
B. japonicum	USDA 6	SAMD00060992	Complete	9.20738	63.7	1	absent
	J5	SAMN05890661	Complete	10.1387	63.3	1	soybean nodulation
	5038	SAMN15394813	Complete	9.22625	63.7	1	Soybean nodulation
	E109	SAMN03262953	Complete	9.22421	63.7	1	Farmland soybean rhizoma
	SEMIA 5079	SAMN02726028	Chromosome	9.58303	63.5	1	Duress soybean plants
	NBRC 14783	SAMD00097546	Contig	9.09656	63.7	177	soybean
	5873	SAMN13738681	Scaffold	9.16023	63.7	141	soybean
	Is-34	SAMN03083461	Scaffold	10.3266	63	248	soybean nodulation

	22	SAMN02441445	Scaffold	7.50456	64.5	4	absent
	in8p8	SAMN02440647	Scaffold	7.58992	63.8	52	absent
	is5	SAMN02440582	Contig	7.58879	63.8	60	absent
	USDA 38	SAMN02440784	Scaffold	9.60897	63.5	107	soybean
	FN1	SAMN02666820	Scaffold	9.1385	63.7	87	soil
	UBMA197	SAMN06077198	Contig	10.4422	63.3	287	Panleaf pinoma rhizoma
	USDA 123	SAMN02441447	Scaffold	10.4577	63.3	517	soybean
	CCBAU 25435	SAMN02469483	Contig	9.46079	63.5	520	absent
	CCBAU 15618	SAMN02469476	Contig	9.82401	63.4	691	absent
	USDA 135	SAMN02441452	Scaffold	7.70332	64	547	soybean
	CCBAU 15354	SAMN02469475	Contig	10.1266	63.3	951	absent
	CCBAU 15517	SAMN02469482	Contig	9.91703	63.4	1129	absent
	CCBAU 83623	SAMN02469465	Contig	10.0743	63.3	1212	absent
B. diazoefficiens	USDA 110	SAMN03573437	Complete	9.10606	64.1	1	absent

2.2 Genomic Integrity

The genomic information obtained is tested with CheckM (PARKS, 2015) and BUSCO (SIMAO, 2015), and the genome with insufficient completeness or high contamination is not conducive to the Pan-genomic subsequent analysis. CheckM evaluates genome integrity and contamination by specific to a species lineage and unique genes in the database. Cds of genome are predicted through Prodigal (HYATT, 2010) software that Prodigal shows excellent robustness to gene structure prediction, translation starting site recognition, and false positives. BUSCO constructed a single-copy conserved gene set of genomome Rhizobiales through the OrthoDB database, compared the transcript results by Augustus software, and then the proportion and integrity to evaluate genome integrity.

2.3 Pan-genome Analysis

Genomes were first annotated through Prokka software, analyzed with gbk files of generated GeneBank through BPGA (CHAUDHARI, 2016) (V1.3), selected default parameters. BPGA adopted Neighbour Joining Tree achievements for core genes and used USEARCH Clustering Algorithm genome annotation is mainly based on COG and KEGG databases. The Pipeline of, panX (DING, 2018), PGCGAP (https://github.com/yikedou/pgcgap).

2.4 Analysis of Secondary Metabolites

The secondary metabolites of microorganisms are a class of complex functional compounds synthesized by primary metabolites through complex synthetic paths and processes, such as antibiotics, pheromones, toxins, etc., which is very important for the growth and competition of microorganisms. The more complex the life history, the wider the host, the greater the living environment changes. The more frequent the communication with the host, the more metabolites of the secondary level. Secondary metabolites can be predicted through published microbial genomic data. Secondary metabolites analysis can be analyzed based on antiSMASH. Still, the online version is more abundant, comprehensive and accurate than the local version, so the online version of antiSMASH (https: / / antismash. secondary etabolites.org), the parameters are default. The strictness of the test is relaxed.

3 RESULTS AND DISCUSSION

3.1 Genomic Integrity

After the genome integrity analysis of 22 soybean slow biological rhizomas, we found that the selected genome integrity was higher, all above 97%, but only a few genomes could reach 100%, as shown in Figure 1. The CheckM results showed that the remaining genome was relatively complete except for a weak B. japonicum CCBAU series sequence deletion (Figure 1), consistent with the BUSCO results (Figure 2).In addition, most of the genome has a small number of

gene pollution, the main reason for this phenomenon may be wrong sequencing, foreign genes (human, bacteria in the air, etc.), pollution or sequence splicing in the sequencing process, but this part is relatively low and does not affect the subsequent analysis.BUSCO analysis showed a small number of genomic and other analysis processes.

genes (Figure 1B yellow section), generally due to insufficient sequencing depth or incomplete genome splicing. In addition, the BUSCO results show multiple copies of parts of the Marker gene in the rhizobium genome. Overall, the NCBI uploads high genome integrity and allows downstream generic

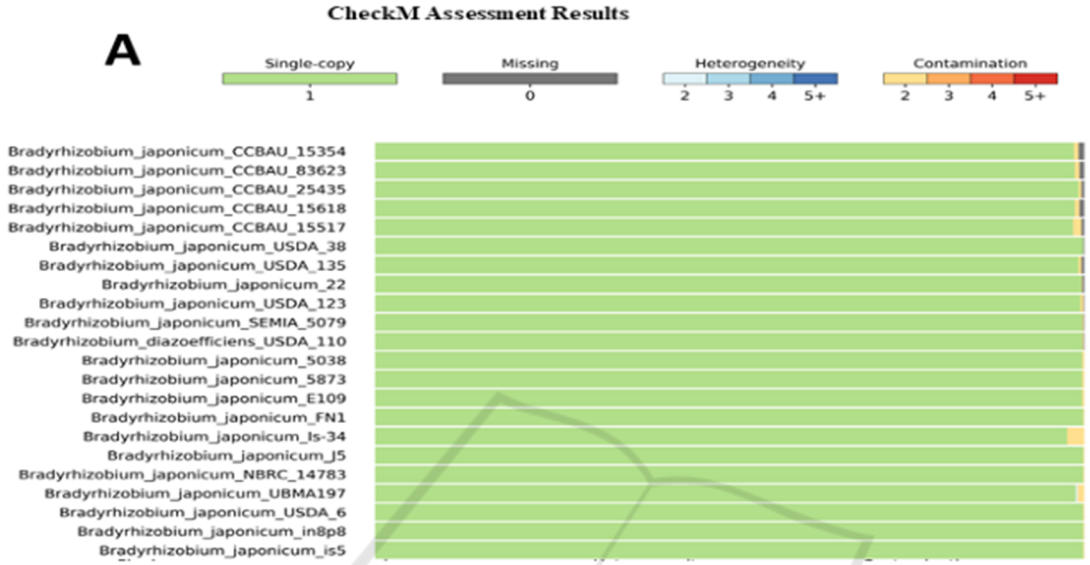


Figure 1: 22 Genome integrity assessment of Bradyrhizobium(checkM Assessment Result).

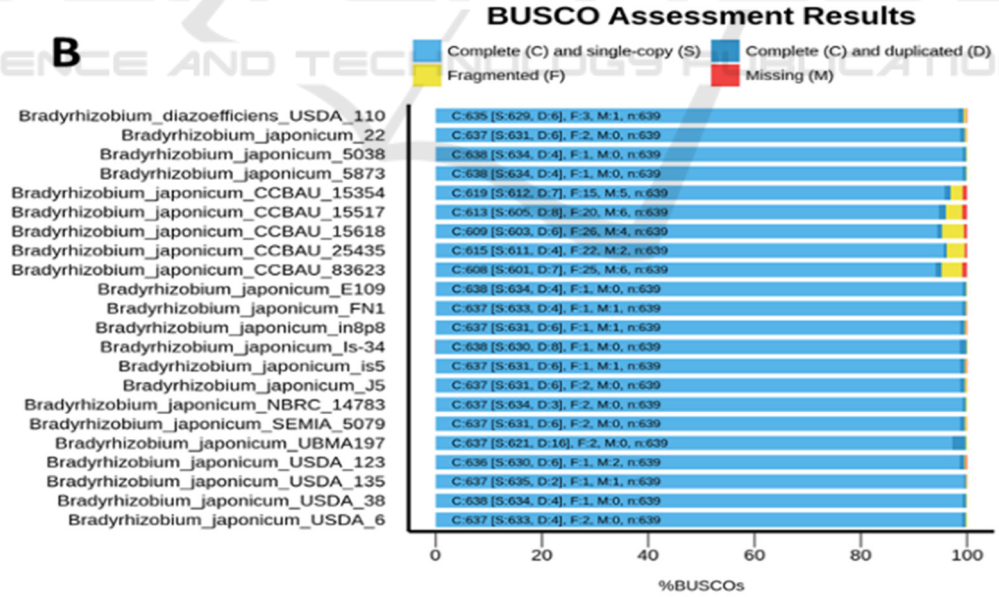


Figure 2: 22 Genome integrity assessment of Bradyrhizobium(BUSCO Assessment result).

3.2 Pan-genomic Statistics

This paper counted the number of core, accessory, and unique genome (specific) and unique deletion genome of 22 Rhizobium strains, as shown in Table 2. Bradyrhizobium diazoefficiens USDA 110 is found more common genomics (794) specific deletion genes (63), so although its genome is complete, it still varies from Bradyrhizobium japonicum, so it is understandable to divide it into other categories. The Pan-genomic statistics of 22 rhizobium counted 3,807 core genes, which is very small. The core genome is the genes shared by all strains involved in basic biological processes such as gene expression, energy production, amino acid metabolism, etc. Some strains have even more accessory genomes than the number of core genomes. Accessory genomes represent some specific functions and have a relatively strong metabolic ability. Unique genomes represent that some of their achievements are more competitive. The more unique genomes, the stronger the resistance to environmental adaptability. Thus it can be seen B.japonicum UBMA197, B. japonicum USDA_135, B. japonicum 22 metabolic ability, environmental adaptability ability are very strong; B.japonicum 5038, B. japonicum 5873, B. japonicum E109, B. japonicum FN1 is highly metabolizing, but it is less heterophenetic and less adaptable. Many genomes have more unique genomes like B. japonicum USDA_135, B. japonicum 22, B. diazoefficiens USDA_110, B. japonicum UBMA197, which make each biometabolic process rich and will be beneficial to the expansion of habitat (Konstantinidis, 2004), which may be the reason why the wide range of Bradyrhizobium adapted to (Tian, 2012).

Table 2: 22 Pangenic statistics of rhizoma strains.

Species	Core genes	accessory gene	unique gene(specific)	Specific deletion gene
B. diazoefficiens USDA_110	3807	3213	794	63
B. japonicum USDA_123	3807	4745	527	12
B. japonicum USDA_135	3807	2242	717	114
B. japonicum 22	3807	2019	758	114
B. japonicum 5038	3807	4155	1	0
B. japonicum 5873	3807	4148	0	1
B. japonicum CCBAU_15354	3807	5010	132	24
B. japonicum CCBAU_15517	3807	4985	170	22
B. japonicum CCBAU_15618	3807	4422	433	22
B. japonicum CCBAU_25435	3807	4122	347	22
B. japonicum CCBAU_83623	3807	5032	266	21
B. japonicum E109	3807	4152	1	0
B. japonicum FN1	3807	4152	6	1
B. japonicum in8p8	3807	2919	4	1
B. japonicum is5	3807	2918	12	1
B. japonicum Is-34	3807	4409	645	0
B. japonicum J5	3807	4359	400	1
B. japonicum NBRC_14783	3807	4135	10	3
B. japonicum SEMIA_5079	3807	4143	246	1
B. japonicum UBMA197	3807	4176	1048	8
B. japonicum USDA_6	3807	4141	17	5
B. japonicum USDA_38	3807	4276	314	1

3.3 COG Analysis

COG analysis found that,as shown in figure 3, the genome was mainly focused on related functions such as [R]General function prediction only, [E]Amino acid transport and metabolism,[K]Transcription. The proportion of the core genome, accessory genome, and unique genome is almost the same in these functions, directly related to biological trait expression and basic functions, so the genome plays a major role in biological trait expression.

The numbers of genomes related with [M] cell wall\ membrane\ evelope biogenesis, [T] singal transduction mechanism, [L] replication, recombination and repair, [C] energy production and conversation, [G] Carbohydrate transport and metabolism, [I] lipid transport and metabolism, [Q] secondary metabolites biogenesis, transport, and catabolism, [P]inorganic ion transport and metabolism are similar. It shows that the strain gene is mostly concentrated in the metabolic process and cell differentiation process. It is preliminarily speculated that this is suitable for the nitrogen fixation process of rhizobium and legume symbiosis. The nitrogen fixation process is very complex. It includes many symbiotic processes, such as rhizoma infection, bacteria-like differentiation, and tuberous nitrogen fixation. Their differentiation and rhizobiums require the expression of related genes in a large number of cell components. The invasion and symbiotic nitrogen fixation process with rhizomes involves biological processes, as well as signal exchange, interaction, material transportation, and metabolism, so it requires the participation of many genes related to molecular function and biological processes (Wang, 2014). The above inferences now require further testing.

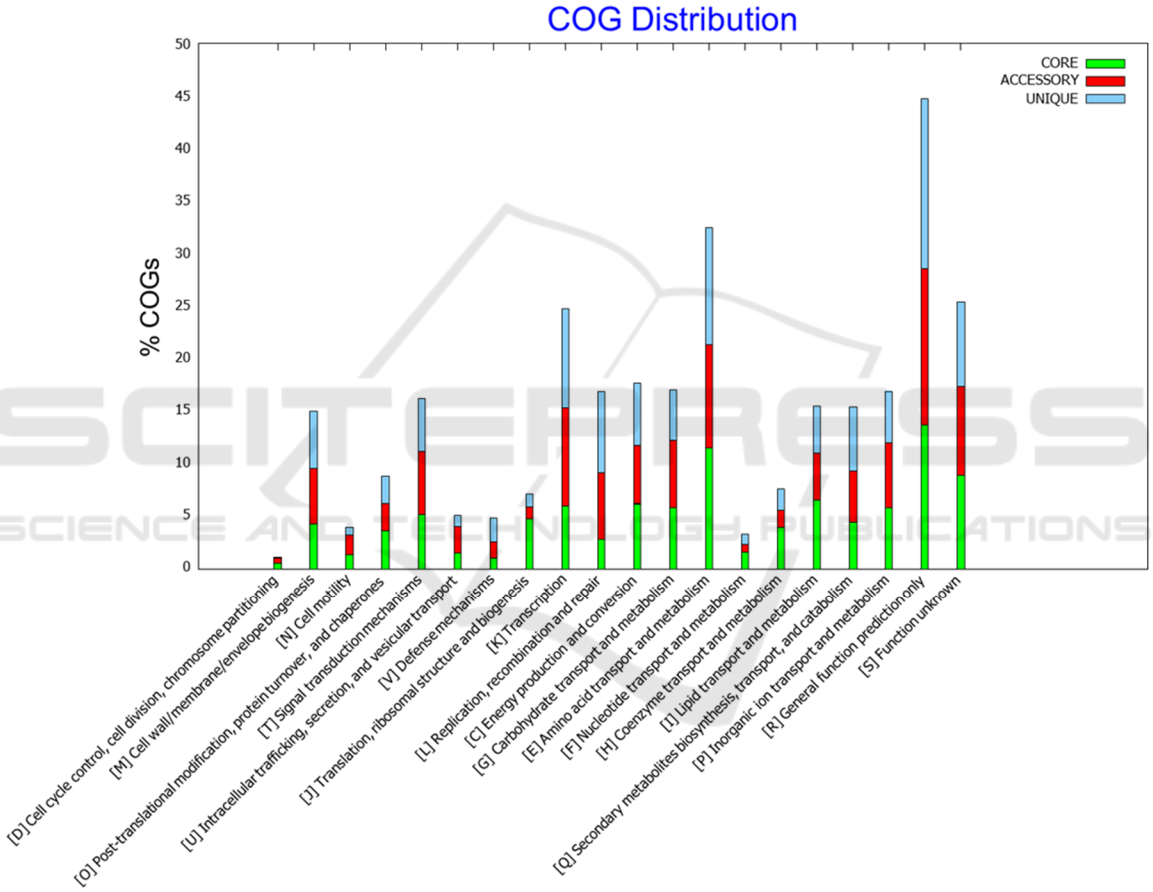


Figure 3: COG analyse.

3.4 KEGG Analysis

As shown in figure 4, KEGG analysis found a high proportion of genome distribution in amino acid metabolism, carbohydrate metabolism, membrane transport, overview, etc. In these fundamental functions, like amino acid metabolism, core genomes play a major role. Amino acid transport and metabolism are relatively many genes, and it is preliminarily speculated that this may be related to its high activity efficiency. Because complex amino acid circulation is essential for symbiotic nitrogen fixation, both sides will control the substance exchange by controlling amino acid metabolism (Lodwig, 2003); (Prell, 2006). At the same time, the genes are distributed in cardiovascular diseases, cellular community, development, digestive system, excretory. The number of specific functions such as

system, sensory system, signaling molecules, and interaction\ substance dependence\ transcription is small. Therefore, it can be seen that these bacteria all have a strong basic function, but the specific environment is poorly adaptable.

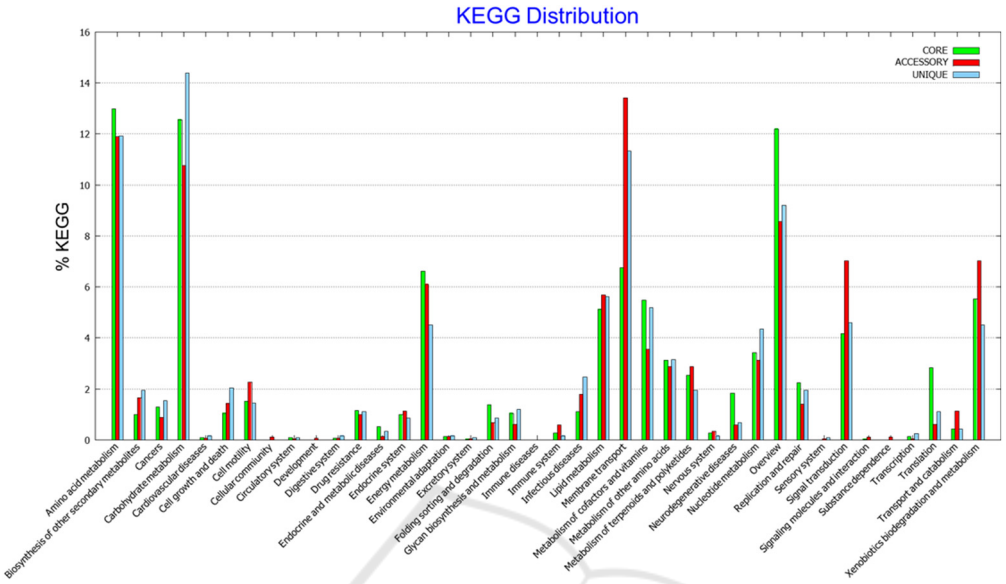
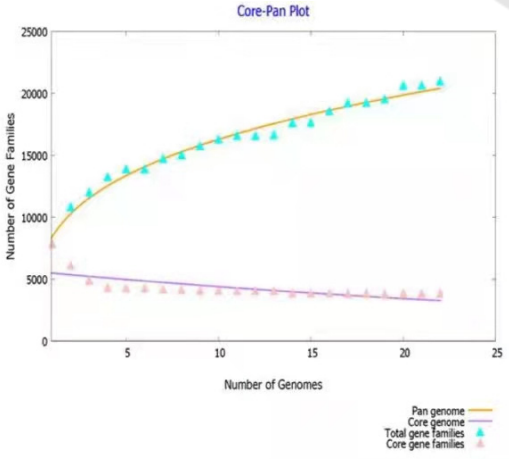


Figure 4. KEGG analyse

3.5 Panogenomic Fitting Equations

The correlation parameters of the fitting equation, as shown in figure 5, of the generic genome size (T) relation to genome number (X) are shown in the figure, showing that the number of genes in the generic genome of Bradyrhizobium increases as the number of the genome increases. Different genomes have the same gene family, and when the number of genomes in each family increases, the genes in this family are relatively open (Sun, 2013). The change is unique number, and the curve shows the core gene number with an increasing genome. Open here means that the species is able to exchange genetic material with other species in many different ways to acquire new genes. The Bradyrhizobium genome has extremely high plasticity, suggesting that it may more readily acquire new genes to accommodate complex changes in the environment.



	Pan-genome	Core-genome
Fit law	POWER	EXPONENTIAL
Equation	f(x)=a.x^b	f1(x)=c.e^(d.x)
Parameters	a= 8216.77	c= 6065.22
	b= 0.301849	d= -0.0305847
Expected Size	20945	0
Estimated Size	20888.64	3094.75

The parameter 'b' = 0.301849 , The pan genome is OPEN.

Figure 5: Panogenomic fitting equations.

3.6 Analysis of Secondary Metabolites

Secondary metabolites are predicted through published microbial genomic data and are analyzed using an online version of antiSMASH (https: / / antismash. secondary etabolites.org). Similar similarities were found between strains, as shown in table 3.

lanthipeptide-class-v, cyanobactin(cyanobacteria), beta lactone, T3PKS, T1PKS, phosphonate, is compounds do not occur in other species. Hserlactone and RiPP-like, terpene, LAP, proteusin all exist to maintain the basic metabolic process. Lipoids such as beta lactone, Hserlactone can serve as stored energy in extreme environments or hungry situations for microbial growth to provide carbon sources and energy (Kadouri, 2005). Compounds such as terpene are present in all types of strains, are associated with the growth and development of plants, and participate in the plant defensive response to (LYU, 2017). Therefore, the enrichment of lipid metabolism and transport pathway and the participation of compounds such as terpene may be mechanisms for the effective environmental adaptation of Bradyrhizobium. Lanthipeptides and RiPP-like is a large class of natural peptide products synthesized by ribosomal and translates modified. Such compounds are widely produced in different bacteria, with rich structural and biological activity diversity; T3PKS and T1PKS all belong to the antimicrobial proteins generated by the ribosomal pathway. RRE-containing rarely occurs, only where Bradyrhizobium japonicum CCBAU 15517, Bradyrhizobium japonicum E109, is endemic to secondary metabolites. RRE-containing is related to RNA transcription and gene expression, so Bradyrhizobium japonicum CCBAU 15517, Bradyrhizobium japonicum E109 will have some specific functions, representing some specific functions its strong environmental adaptability.

Table 3: Analysis of the secondary metabolites of 22 rhizoma strains.

Baterial name	Number of secondary metabolites gene clusters	secondary metabolite
Bradyrhizobium japonicum 22	11	hserlactone, RiPP-like, T1PKS, terpene, **lanthipeptide-class-v**, **cyanobactin**, redox-cofactor, **beta lactone**
Bradyrhizobium japonicum 5038	10	RiPP-like, proteusin, NRPS, ectoine, NRPS, terpene, LAP, redox-cofactor, hserlactone
Bradyrhizobium japonicum 5873	10	hserlactone, RiPP-like, NRPS, redox-cofactor, LAP, proteusin, terpene
Bradyrhizobium japonicum CCBAU 15354	15	**T3PKS**, hserlactone, RiPP-like, terpene, NRPS, redox-cofactor, LAP, **lanthipeptide-class-v**, RiPP-like
Bradyrhizobium japonicum CCBAU 15517	13	RRE-containing, hserlactone, NRPS, RiPP-like, LAP, terpene, T3PKS, proteusin
Bradyrhizobium japonicum CCBAU 15618	14	hserlactone, terpene, LAP, redox-cofactor, proteusin, NRPS, NRPS-like
Bradyrhizobium japonicum CCBAU 25435	13	terpene, LAP, hserlactone, redox-cofactor, RiPP-like NRPS, proteusin
Bradyrhizobium japonicum CCBAU 83623	14	redox-cofactor, hserlactone, NRPS, terpene, RiPP-like, LAP, T3PKS, ectoine, proteusin
Bradyrhizobium japonicum E109	10	hserlactone, redox-cofactor, terpene, LAP, RiPP-like, NRPS, ectoine, proteusin, RRE-containing,
Bradyrhizobium japonicum FN1	10	hserlactone, redox-cofactor, LAP, terpene, NRPS, RiPP-like, proteusin
Bradyrhizobium japonicum in8p8	9	hserlactone, redox-cofactor, terpene, T1PKS, RiPP-like, NRPS, beta lactone
Bradyrhizobium japonicum is5	9	hserlactone, redox-cofactor, terpene, T1PKS, RiPP-like, beta lactone
Bradyrhizobium japonicum Is-34	12	redox-cofactor, NRPS, LAP, phosphonate, hserlactone, NRPS, terpene, proteusin, hserlactone, RiPP-like
Bradyrhizobium japonicum J5	15	hserlactone, redox-cofactor, RiPP-like, terpene, LAP, NRPS, ectoine, proteusin, phosphonate, T3PKS
Bradyrhizobium japonicum NBRC 14783	10	LAP, terpene, hserlactone, RiPP-like, NRPS, redox-cofactor, proteusin

Bradyrhizobium japonicum SEMIA 5079	11	RiPP-like, proteusin, ectoine, NRPS, terpene, LAP, T1PKS, redox-cofactor, hserlactone
Bradyrhizobium japonicum UBMA197	17	hserlactone, phosphonate, proteusin, LAP, RiPP-like, NRPS, T3PKS, terpene, redox-cofactor
Bradyrhizobium japonicum USDA 6	10	hserlactone, redox-cofactor, terpene, LAP, RiPP-like, NRPS, ectoine, proteusin
Bradyrhizobium japonicum USDA 38	12	LAP, RiPP-like, NRPS, hserlactone, proteusin, terpene, redox-cofactor
Bradyrhizobium japonicum USDA 123	20	terpene, redox-cofactor, LAP, hserlactone, T3PKS, NRPS, RiPP-like, T1PKS, proteusin, ectoine
Bradyrhizobium japonicum USDA 135	10	hserlactone, RiPP-like, redox-cofactor, NRPS, ectoine, terpene, NRPS
Bradyrhizobium diazoefficiens USDA 110	9	hserlactone, proteusi, RiPP-like, NRPS, ectoine, terpene, NRPS-like, LAP, redox-cofactor

4 CONCLUSION

This paper systematically studied the genome sequence of 22 Bradyrhizobium strains, which finds that the genome size is within the range of 7.50456Mb-10.4577Mb, and the selected genome integrity is high, all above 97%. The genomes all have 3807 core genes, with an open genome.COG analysis found that the Pseudomonas genome had a higher proportion of genes related to the underlying metabolic functions such as General function prediction only, Amino acid transport and metabolism, Transcription. Analysis of the secondary metabolites found that most of the secondary metabolites of the strain were T3PKS, peptides, terpene, and esters. However, due to the length and the small number of reference whole genomes, the separation environment, and evolutionary relationship still need to be strengthened. In later work, broader Bradyrhizobium strains can be collected, with more systematic and in-depth research on the relationship between evolutionary history and environmental adaptation, evolutionary environment, and genomic characteristics.

REFERENCES

BROGHAMMER A, KRUSELL L, BLAISE M, et al. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding[J]. PNAS, 109(34): 13859-13864 (2012).

BARSY M D, FRANDI A, PANIS G L, et al. Regulatory (pan-)genome of an obligate intracellular pathogen in the PVC superphylum[J]. The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 10(Pt 2)2016.

CHAUDHARI N M, GUPTA V K, DUTTA C. BPGA- an ultra-fast pan-genome analysis pipeline [J]. Scientific Reports, 6 (2016).

DING W, BAUMDICKER F, NEHER R A. panX: pan-genome analysis and exploration [J]. Nucleic Acids Research, 46(1) (2018).

Herridge, D.F., Peoples, M.B. & Boddey, R.M. Plant Soil 311, 1–18 (2008).

HYATT D, CHEN G-L, LOCASCIO P F, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification [J]. Bmc Bioinformatics, 11 (2010).

Konstantinidis KT, Tiedje JM. Trends between gene content and genome size in prokaryotic species with larger genomes [J]. PNAS, 101 (9): 3160-3165 (2004).

Kadouri D, Jurkevitch E, Okon Y, Castro-Sowinski S. Ecological and agricultural significance of bacterial polyhydroxyalkanoates[J]. Crit Rev Microbiol, 31 (2): 55-67 (2005)

L R, V M, P-E F, et al. The bacterial pangenome as a new tool for analysing pathogenic bacteria[J]. New microbes and new infections, 7 (2015).

L M, A M, B D, et al. Nodulation of legumes by members of the beta-subclass of Proteobacteria[J]. Nature, 411(6840) (2001),.

Lodwig EM, Hosie AH, Bourdes A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS. Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis [J]. Nature, 422 (6933): 722-726 (2003).

LYU HZ, LIU WJ, HEL, etal. Advances on the study of gene clusters involved in plant secondary metabolism [J]. Plant science journal, (in Chinese with English abstract) ,35 (4):609G610,612G621(2017)

MATT L, S H P, F W M. Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary [J]. Narnia, 54(4) (2005).

PARKS D H, IMELFORT M, SKENNERTON C T, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes[J]. Genome Research, 25(7): 1043-1055 (2015).

Prell J, Poole P. Metabolic changes of rhizobia in legume nodules [J]. Trends Microbiol, 14 (4): 161-168 (2006)

Stoksta E. Science, 353(6305): 1225-1227 (2016).

SPRENT J I, ARDLEY J, JAMES E K. Biogeography of nodulated legumes and their nitrogen-fixing symbionts[J]. New Phytologist, 215(1): 40-56 (2017) .

SIMAO F A, WATERHOUSE R M, IOANNIDIS P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs [J]. Bioinformatics, 31(19): 3210-3212 (2015).

Sun C. Research and development of comparative genomics visualization system (CGVS) [D]. Beijing: The University of Chinese Academy of Sciences, (2013).

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR, Chen WX. Comparative genomics of rhizobianodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations [J]. PNAS,109(22):8629-8634 (2012).

Wang SM. Whole-genome sequencing and comparative genomic analysis of Mesorhizobium huakuii strain 7653R [D]. Wuhan: Huazhong Agriculture University (2014).