Identification of gene orthologs and molecular markers for drought-tolerant nitrogen-fixing *Bradyrhizobium* strains isolated from soil in Texas

Mohammad Kamruzzaman*, and Woo-Suk Chang
Department of Biology, The University of Texas Arlington, TX 76019

Abstract

Bradyrhizobium japonicum is a soil bacterium capable of establishing a symbiotic relationship with legume plants such as soybean. This symbiosis leads to the formation of root nodules in which the bacteria convert atmospheric nitrogen into ammonia, an essential nutrient source for the plant. However, this mutualistic association is significantly impacted by drought, a prominent abiotic stressor in US soybean cultivation. Thus, developing drought-tolerant soybean cultivars and compatible rhizobial strains has become a growing concern in the US agricultural industry. In a previous study, we identified two drought-tolerant *Bradyrhizobium* strains, namely TXVA and TXEA, isolated from Texas soils. Subsequently, we sequenced the complete genomes of both strains, revealing that TXVA's genome comprises 9,193,770 base pairs with 8,980 protein-coding genes, while TXEA's genome contains 9,339,455 base pairs with 9,158 protein-coding genes. In this study, we employed a computational strategy to identify genes with analogous functions across different strains. Using OrthoFinder, we found genes exhibiting comparable activities in TXVA, TXEA, and seven other known rhizobial strains. Furthermore, employing the Pangenome pipeline, we identified unique gene clusters within all compared genomes. Additionally, the Phylogenetic Profiler for Single Genes (IMG, JGI) enabled us to pinpoint distinctive genes for both TXVA and TXEA. The pangenome analysis revealed that a gene encoding a hypothetical protein (IMG gene ID: 2929624621) and another gene encoding deoxycytidylate deaminase were identified as unique to TXVA and TXEA, respectively.

Introduction

TXVA and TXEA are two *Bradyrhizobium* strains isolated from Victoria and Lubbock counties in Texas. These two strains showed better desiccation survivability compared to *Bradyrhizobium japonicum* USDA110 in our previous study. According to our recent field study data, these strains have shown better symbiosis performance and an overall increase in soybean yield over commercially available soybean inoculants. The genome sequence of these two strains has been done recently, and due to its overall advantage as a soybean inoculant, these two strains got our attention for the genome comparison study to identify functionally similar genes and unique molecular marker genes in these two native rhizobia isolates.

Materials and Methods

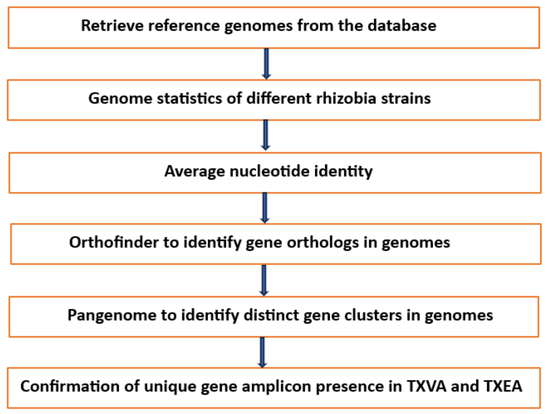

Figure 1. Schematic outline of working pipeline. Here we compared, *Bradyrhizobium japonicum* TXEA (GCA_023511055.1), TXVA (GCA_023511065.1), USDA110 (BA000040.2), SEMIA5079 (CP007569.1), E109 (CP010313.1), USDA6 (AP012206.1), *Sinorhizobium. meliloti* Rm1021(AL591688.1), *Rhizobium etli* CFN42 (CP000133.1), *Rhizobium leguminosarum bv trifolii* CC275e (CP053439.1). Strains were selected based on agronomic use and desiccation tolerance. Protein sequence (.faa) was used for Orthofinder comparison, and Nucleotide sequence was used for pangenome analysis, ANI, and genome statistics.

Results

1. Genome statistics: This was done to identify major differences between all compared genomes. The genome statistics data shows that among these nine rhizobia strains, the total number of genes was highest in TXEA (9277) then TXVA (9098); while model rhizobia USDA110 has (8402) and lowest observed in CFN42 (6117). GC% was highest in USDA110 (64.06) and found lowest in CC275e (60.88).

2. Average Nucleotide Identity: ANI was done to identify mostly similar and dissimilar Rhizobia against TXVA and TXEA. According to ANI, TXVA and TXEA both are similar to E109 (TXVA ANI-98.98, AF-97.29; TXEA ANI-98.98, AF-96.90) and dissimilar to CC275e (TXVA ANI-74.64, AF-8.43; TXEA ANI-74.63, AF-8.58). TXEA and TXVA also have a good similarity to each other (ANI-98.98, AF-96.60).

3. Gene orthologs: Orthofinder pipeline was followed to identify functional and structurally similar genes. The output was generated as a . CSV file for all gene orthologs from all compared genomes. Here, we present a sample table for desiccation-related genes; from this table, we can easily locate similar genes for TXVA and TXEA.

Gene name	orthogroup	*Bradyrhizobium japonicum* USDA110	Bradyrhizobium japonicum E109	*Bradyrhizobium japonicum* SEMIA 5079	TXEA	TXVA
otsB (trehalose phosphatase)	OG0002720	BAC45588.1 (otsB) 273 aa	AJA59230.1 (trehalose phosphatase) 273 aa	AHY56012.1 (trehalose-phosphatase) 273 aa	2929630997 (trehalose 6-phosphate phosphatase) 273 aa	2929618920 (trehalose 6-phosphate phosphatase) 273 aa
treT (trehalose glycosyl transferring synthase)	OG0000032	BAC49185.1, BAC52331.1	AJA60934.1, AJA63857.1, AJA66124.1	AHY51274.1, AHY54193.1, AHY56455.1	2929628415, 2929629878, 292963 0744, 2929632610	2929619090, 2929621815, 29296233 00

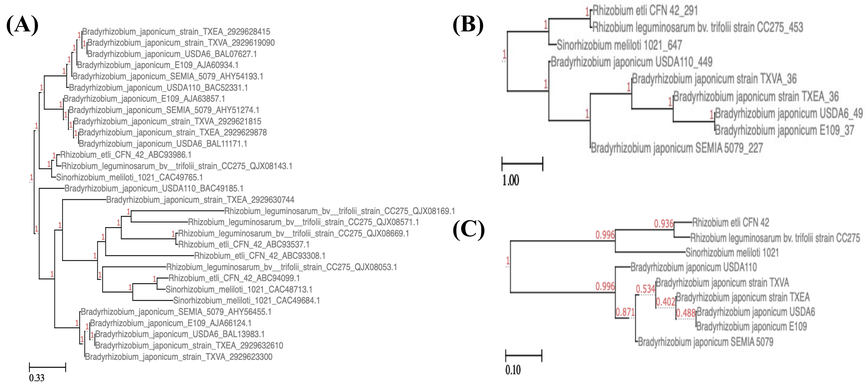

Figure 2. Gene orthologs and relationship of genes of TXVA, TXEA with other rhizobia strains. (A) Rooted gene phylogenetic tree, (B) Gene duplication events, (C) Rooted species tree.

4. Distinct gene identification by Pangenome analysis: The Pangenome is a useful genome comparison pipeline to compare population structure, identify common and unique genes in a genome, and identify any evolutionary changes in the genomes.

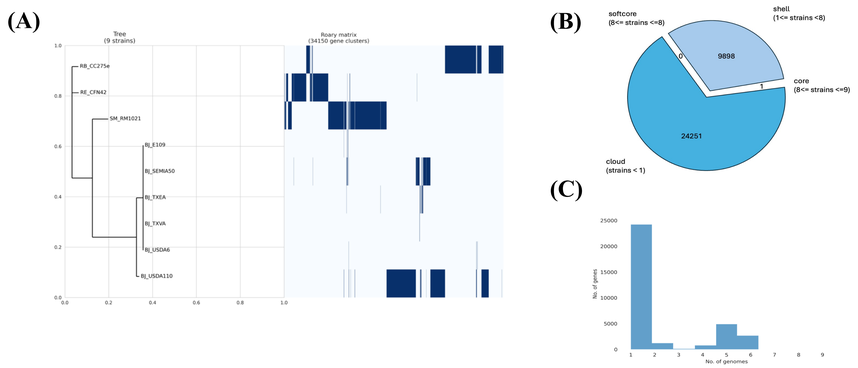

Figure 3. (A) Heatmap of the entire pangenome matrix demonstrating which gene clusters are shared by the genomes and which genes are exclusive to each strain. **(B)** The Pangenome Pie chart. **(C)** Pangenome frequency plot.

5. Blast to validate absence in any other organism: 28 and 217 unique genes were found for TXVA and TXEA by Roary (pangenome) analysis, and IMG-JGI tool (Phylogenetic profilers for the single gene) could identify 43 and 153 genes respectively for TXVA and TXEA. An NCBI nucleotide blast was carried out to crosscheck the presence of these unique genes in any other organism in the database.

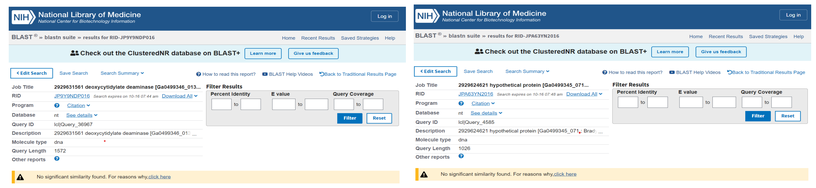

Figure 4. NCBI nucleotide blast represents no match for these two genes.

6. Unique gene confirmation using PCR: To check the presence of the gene, we used PCR to amplify the gene from TXEA and TXVA. PCR confirmed TXEA Deoxycytidylate deaminase gene, and TXVA hypothetical protein gene amplification is still ongoing.

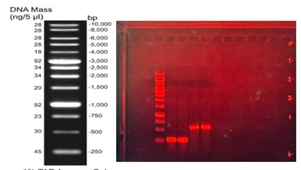

Figure 5. PCR confirmed deoxycytidylate deaminase gene presence in TXEA. Here, lane 1 is a 1kb ladder, lane 2-3 is the housekeeping gene parA-partial amplicon (287/855bp) as a control, and lane 4-5 is the deoxycytidylate deaminase partial gene amplicon (577/1572bp).

Conclusions

- ***rpsJ*** was found as a core gene in all compared genomes. This 30S ribosomal protein S10 functions in translation, ribosomal structure, and biogenesis.
- Although TXVA and TXEA have 43 and 153 unique genes, most of which match *Burkholderia* species, a gram-negative soil bacteria.
- Deoxycytidylate deaminase (catalyzes the conversion of dCMP to dUMP in the deamination process) found unique in TXEA.
- A hypothetical protein (IMG gene ID: 2929624621) was found unique for TXVA.

Future Directions

- Distinct genes identified from TXEA and TXVA can be used as molecular markers and can be used for field identification of these two strains.
- Functional identification and validation of these two genes can be done.
- Orthofinder data can be used to select similarly functioning genes in these two native rhizobia isolates.

References

1. Emms, D.M. and Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology, 20, pp.1-14.
2. Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics, 31(22), pp.3691-3693.
3. Peterson, C., Niraula, S., Parks, D. and Chang, W.S., 2022. Draft Genome Sequences of Two Desiccation-Tolerant Strains, Bradyrhizobium japonicum TXVA, and TXEA, Isolated from the Root Nodules of Soybean Grown in Texas. Microbiology Resource Announcements, 11(8), pp.e00467-22.