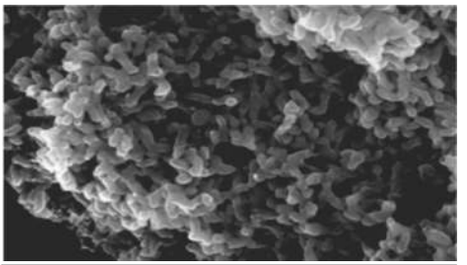
AMAAS

Application of Microorganisms in Agriculture and Allied Sectors



Indian Council of Agricultural Research (ICAR), New Delhi started a National Network Research Project on "Application of Microorganisms in Agricultural and Allied Sectors (AMAAS)" in the 10th five year plan. Under the AMAAS project *Mesorhizobium ciceri* Ca181 was selected for complete genome sequencing.

चना राइजोबियम का अनुक्रमण

Decoding of the Chickpea Rhizobium Genome

First complete microbial genome sequence from India

- The current concept of the bacterial species – a 'genomically coherent' group of strains that share many common traits rests squarely on the availability of reliable techniques to quantify the relatedness of bacterial genomes. An alternative approach to quantification of genome relatedness is to compare selected DNA sequences for a group of bacterial strains. The core technology for this method, DNA sequencing, is relatively rapid, inexpensive and highly reproducible. Databases of gene sequences and computer applications to compare them are, likewise, freely available. For these reasons, several projects on whole genome sequence analysis has been undertaken by the different groups of researchers.
- The field of microbial genomics has moved away from the primary initial focus on pathogens genomes to include the sequencing of diverse prokaryotes that occupy a range of environmental niches, and which are responsible for an array of environmental processes. Every genome that has been sequenced to date has provided new insight into biological processes, activities, and potential of these species that was not evident before the availability of the genome sequence.
- The decoding of genomes is crucial for a rapid advancement in genetic improvement of bio-fertilizer and bio-control agents, bio-energy organisms, varieties of agricultural plants and animal breeds. The countries leading the global genomics research are now contemplating sequencing of hundreds and thousands of genomes which has now become possible due to infusion of second and third generation of sequencing technologies and development of computational biology.

The dedicated team of AMAAS researchers has now sequenced *Mesorhizobium ciceri* Ca181; the first whole genome sequence of one of most agriculturally important bacterium from India. This is a landmark achievement in the frontier area of biological research that involves mastering of latest techniques in genome sequence assembly and annotation. The estimated genome size of *M. ciceri* Ca181 is 6.47 Mbp and it has 6742 predicted genes in 4116 transcription units. Several of these identified genes are involved in biological nitrogen fixation (34 genes) and stress tolerance (184 genes). This discovery may open up the doors for intensified research on Rhizobium strain improvement for bio-fertilizer applications. The information could also be utilized for the preparation of genetically modified improved rhizobium strains that might increase productivity and yield in pulses.

Background

The genomic revolution of the 1990s has resulted in almost 1537 completely sequenced microbial genomes and sequencing of 4531 is in progress. More recently, explosion of random community genomics, or metagenomics, where DNA is sequenced directly from environmental samples has provided insights to understand the nature of microbial communities with the help of 335 completed metagenome sequencing projects. One hundred forty five genomes of agriculturally important microbes have been completely sequenced, and sequencing of more than thousands are in progress. These include projects on sequencing of *Bacillus*, *Pseudomonas*, *Rhizobium*, *Burkholderia*, Cyanobacteria, Fungi and Actinobacteria. These microbial genomes could be utilised for the betterment of agriculture in terms of sustainability, increase in yield, and disease management.

In the traditional genome sequencing approach, the chromosomal DNA is first cloned into Bacterial Artificial Chromosomes (BACs), or small plasmids, and di-deoxy chain termination sequencing (Sanger sequencing) is used to determine the sequences. In the alternative approach used here, the chromosomal DNA was sequenced without cloning, using next-generation sequencing techniques, particularly 454-pyrosequencing and Illumina Solexa. Pyrosequencing has much higher throughput, larger read length and a lower error rate per base sequenced compared to Sanger sequencing.

The project is registered with international registration authorities (Table 1). Internationally only 12 Rhizobium strains have been sequenced and now 13th one i.e. *M. ciceri* Ca181 sequenced which is the first complete agriculturally important microbial genome sequence to be released from India.

The completion of genome sequencing of *M. ciceri* Ca181 has unveiled the specific properties encoded by the genes that work in a coordinated manner to control specific metabolic pathways. After the functional validation of different gene clusters predicted by annotation of genome, we will understand the reason of uniqueness of this bacterium, resulting in discovery of several new genes, operons and proteins.

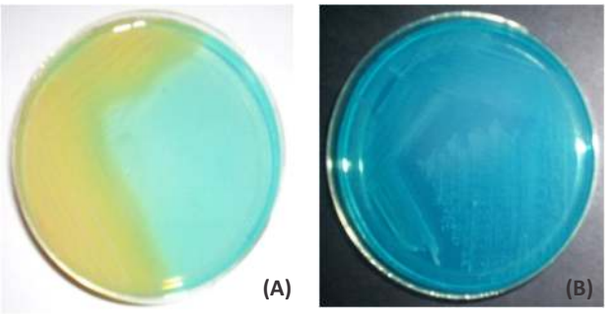


Fig. 1 Ketolactose test; Plates showing colour change of *Agrobacterium* (A) on Ketolactose medium after addition of Benedict's reagent whereas *M. ciceri* Ca181 (B) did not change.

Mesorhizobium ciceri Ca181

M. ciceri Ca181 was isolated from the chickpea root nodules from the fields of Haryana. It is an indigenous strain of India. This bacterium was selected for genome sequencing because it has efficient nitrogen fixation ability, showed good nodulation and significant high competitiveness among other chickpea Rhizobia, and significantly enhance the chickpea yield. This strain has been tested in all India coordinated trials at different locations in different agro-climatic regions and soil types .

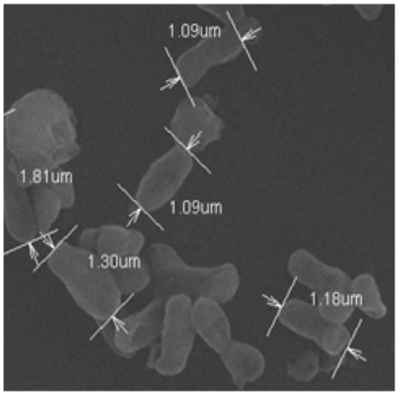
M. ciceri Ca181 is a gram negative, non-spore forming, motile, rod shaped aerobic bacterium. Colonies of *M. ciceri* Ca181 are translucent, milky white as seen after 6-8 days of growth on Yeast Extract Mannitol Agar bacterial medium at 28°C with a generation time of 8h.

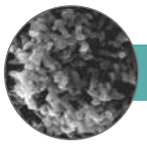
Chickpea Rhizobium (*M. ciceri*) is estimated to add 24-40 kg of nitrogen per hectare in a cropping season. Rhizobium cultures are used as biofertilizer in the legume crops but there is a need to develop improved Rhizobium strains that may fix more nitrogen and could be more competitive when added to the soil as biofertilizer. Knowledge of complete genome structure of the Rhizobium will help to develop improved bio-fertilizer strains.

Registration of the Genome Sequencing Project (NCBI)

Organism name	*Mesorhizobium ciceri*
Strain	Ca181
NCBI Genome Project ID	40923
Locus Tag	M1C
Contact person	Prof. D. K. Arora (Director NBAIM)
Submitting Organization	NBAIM, Mau, India
Chromosome	One
Genome Size	8.0 Mb
URL	http://www.nbaim.org.in
Sequencing centre	NBAIM/NRCPB /NBPGR/IIVR
Consortium Name	AMAAS
Sequencing Technology	ABI Sanger/454-FLX Pyroseuencing/Solexa-illumina
Sequencing Depth	20X

Table. 1 Details of *M. ciceri* Ca181 Genome Sequencing Project Registration at National Centre for Biotechnology Information, U.S.A





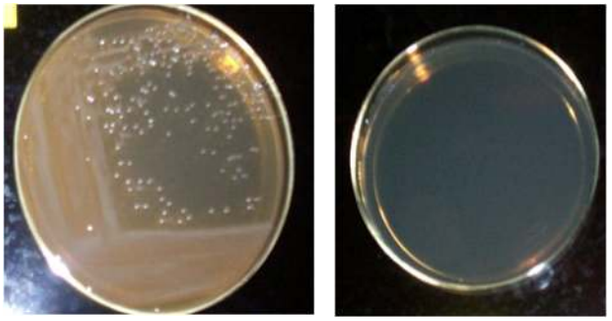


Fig. 2 Plates showing growth of *Agrobacterium* (A) on Hoeffer's medium whereas *M. ciceri* Ca181 (B) did not show any growth

Table.2 Carbon utilization pattern of *M. ciceri* Ca181

S. No.	Carbohydrate source	Result	S.No.	Carbohydrate source	Result
1	Lactose	-	18	Dulcitol	-
2	Xylose	+	19	Inositol	-
3	Maltose	-	20	Sorbitol	-
4	Fructose	-	21	Mannitol	-
5	Dextrose	+	22	Adonitol	-
6	Galactose	+	23	α-methyl-D Glucoside	-
7	Raffinase	-	24	Ribose	+
8	Trehalose	-	25	Rhamnose	+
9	Melibiose	+	26	Cellobiose	+
10	Sucrose	-	27	Melezitose	-
11	L Arabinose	+	28	α-methyl-D Mannoside	-
12	Mannose	+	29	Xylitol	-
13	Inulin	-	30	ONPG	-
14	Sodium gluconate	-	31	Esculin	+
15	Glycerol	-	32	D Arabinose	+
16	Salicin	+	33	Citrate	-
17	Glucosamine	-	34	Malonate	-
18	Dulcitol	-	35	Sorbose	-

Primary check of the authenticity of *M. ciceri* Ca181 culture was done using ketolactose test (Fig. 1) and growth on Hoeffer's alkaline medium (Fig. 2). The bacterium was also screened for utilization of different carbohydrates (Table 2) and tolerance to different concentrations of sodium chloride. *M. ciceri* Ca181 did not have ability to solubilize the soil phosphate was also screened against various antibiotics.

Nodulation and Acetylene Reduction Assay

The nodulation ability of *M. ciceri* Ca181 on chickpea (*Cicer arientinum var.* BG256) was tested along with few other selected strains. Nitrogen fixation activities of root nodules were assayed by the acetylene reduction technique. Sections of nodulated roots were incubated in serum bottles in air with 0.1 atmosphere of acetylene. The amount of ethylene produced was measured by gas chromatography. After the assay, nodules were detached from the root and weighed. Acetylene reduction activity was calculated (Table 3).

Isolate	Fresh Nodule weight (gm)	Nitrogenase activity	No. of pods	Fresh seed weight (gm)	No. of flowers	Fresh Shoot weight (gm)
***M. ciceri* Ca181**	0.818 0.818 0.487	1941.626 743.990 105.53	2 5 9	0.031 0.786 0.457	- 3 -	3.009 3.727 4.870
***M. ciceri* TAL620**	0.218 0.098 0.341	1310.9 73.50 308.9543	1 3 12	0.037 0.136 0.470	3 1 2	5.213 4.072 7.728
***M. ciceri* CH1**	1.045 0.490 0.622	1079.990 50.66 77.2186	4 10 5	0.632 1.020 0.526	1 6 -	5.956 6.897 3.413
***M. ciceri* KT1001**	0.480 0.480 0.477	1011.8208 629.07 308.95	5 4 5	0.226 0.211 0.705	7 4 3	5.093 6.188 4.685
***M. ciceri* KT1002**	1.050 0.818	1932.276 4202.58	13 5	2.503 0.263	3 4	7.852 5.247

Table.3 Acetylene Reduction activity or Nitrogenase activity study of *M. ciceri* Ca181 with other chickpea rhizobium strains.

16S rDNA amplification and sequencing

The total chromosomal DNA of *M. ciceri* Ca181 was isolated and PCR was performed using the primer pair 5'-AGAGTTTGATCCTGGCTAG and 5'-AAGGAGGTGATCCAGCCGCA-3. The products were run on agarose gel and documented before sequencing.

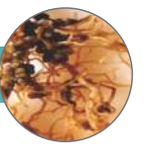
Amplified product was gel eluted and purified for sequencing. DNA sequences of the PCR products were determined using Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, USA). The products of the sequencing reaction were analyzed using a DNA sequencer (3130xl Genetic Analyzer; Applied Biosystems). The BLAST analysis was undertaken using the available database of NCBI GenBank (http://www.ncbi.nlm.nih.gov/blast).

The Chickpea Genome Sequencing Consortium

The genome sequencing work was completed by a consortium of four ICAR institutes: (i) National Bureau of Agriculturally Important Microorganisms, Mau; (ii) Indian Institute of Vegetable Research, Varanasi, (iii) National Research Centre on DNA Fingerprinting, New Delhi and (iv) National Research Centre on Plant Biotechnology, New Delhi, under the network project on "Application of Microorganisms in Agriculture and Allied Sectors", which was overall coordinated by the NBAIM.

Sequencing

The consortium adopted a whole genome shotgun sequencing approach for the decoding of *M. ciceri* Ca181 genome. The traditional Sanger's chain termination sequencing (ABI platform), as well as the next generation sequencing (NGS) technologies including 454-pyrosequencing and Solexa-illumina were used.



ABI Sanger Sequencing

Whole genome shotgun plasmid library as well as large fragment fosmid library was prepared. Both small insert (2-4 kbp) plasmid library (Fig.3) and large insert (35-42 kbp) fosmid library (Fig.5) were sequenced with Sanger sequencing. A total of 11310 transformed small insert plasmid clones and 19x96 fosmid clones with large fragments (35-42 kbp) were prepared. Total 9600 reads were obtained from 4800 short fragment insert clones sequencing by Sanger method. These sequence reads were having paired ends of good quality. After Sanger sequencing 4000 kb sequence was generated (Fig. 8).

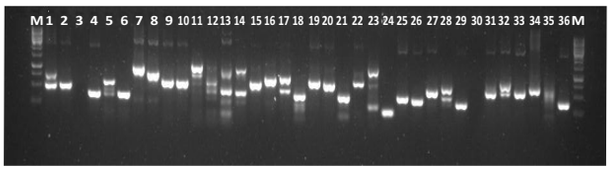


Fig. 3 Amplification of short fragment inserts with M13 consensus primes; Lane M: 1 Kb DNA Marker, Lane 1-36 is amplified product from *M. ciceri* Ca181

454-FLX Pyrosequencing

About 3.7 million reads (12 billion bases) of raw sequence reads were obtained by complete run of 3.5 picotitre plates of FLX-454-pyrosequencing technology with quality value from range 40-64 (Fig 4). FLX library contains fragment size of 400 to 800 bp. Sequence data generated from 454 sequencing was assembled by Newbler Assembler in 47 large sequence contigs (Fig. 7) with the estimated genome size of 6.4Mb and 202X depth coverage of genome (Table 5).

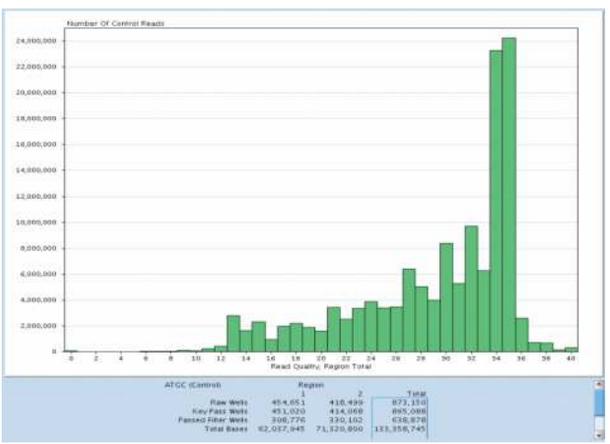


Fig 4 Quality value bars of 454 - FLX Pyrosequencing reads

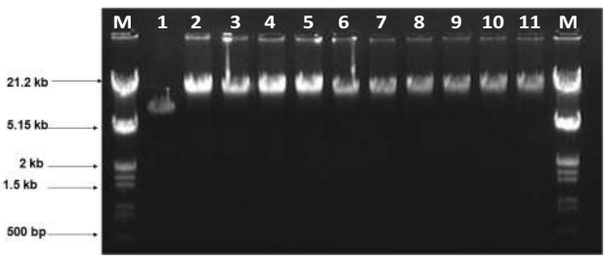


Fig. 5 Lane M - λ *Eco*RI/*Hin*dIII marker, Lane 1 - pCC2FOS cloning vector (8.2 kb), Lane 2-11 - Fosmid DNA isolated from randomly picked fosmid clones of *M. ciceri* Ca181.

Solexa-Illumina Sequence

A single run of Solexa sequence data were assembled using Velvet software. The Velvet software (Table 4) uses de-Bruijn graph algorithms, which is considered to be most effective algorithm for *de-novo* assembly. However, to be on the safe side, we chose to use Velvet 0.7; as it is stable; whereas, Velvet 1.0 has been modified to include comparison with a reference genome as well for reference assembly and unlikely to be stable.

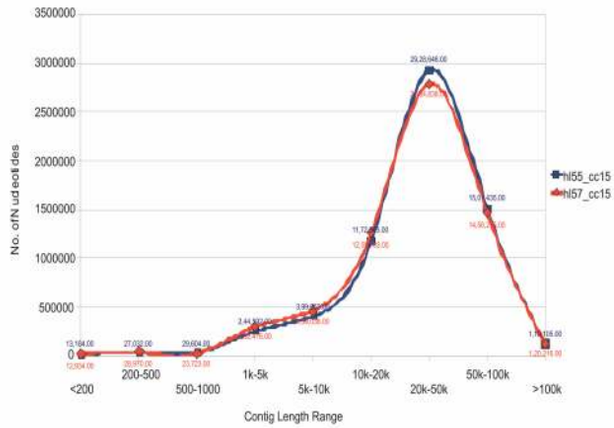


Fig. 6 Contigs coverage and length range from the assembly of Solexa-illumina sequencing data by Velvet Assembler

No of input Sequences	13477268 bp
No of contigs in assembly	562
No of nucleotides assembled	6437289 bp
Minimum contig size	109 bp
Maximum contig size	119105 bp

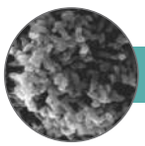
Table 4 Assembly statistics of Solexa illumina sequencing data of ABI Sanger Capillary Sequencing

Genome Assembly

About 12 billion sequences have been generated by 454-pyrosequencing, 1.2 GB by Solexa-illumina and 4000 Kb by Sanger Sequencing. After assembly and merging of 454 and Solexa sequence data total 23 large contigs were made and further assembly using fosmid end sequence data resulted in nine scaffolds having 16 contigs, and 7 contigs are remained unassembled (Table 8). These 23 contigs make up total 6.79 Mbp of finished genome sequence. The scaffold sequences were annotated by different methods available for Genome and genes annotation.

Total number of bases	1276598312 (12 Billion b)
Number of aligned bases	1276598312 (12 Billion b)
Number of Aligned reads	3742907(3.7 Million)
Number of assembled reads	3714101
Total Number of contigs	52
Large contigs	47
Largest contig size	757621 (0.7 Mb)
Average contig size	137246 (0.13Mb)
Genome size	6471179 (6.4 Mb)
Coverage Depth	202X

Table 5. Statistics of the Assembly from 454 sequencing data of *M. ciceri* Ca181



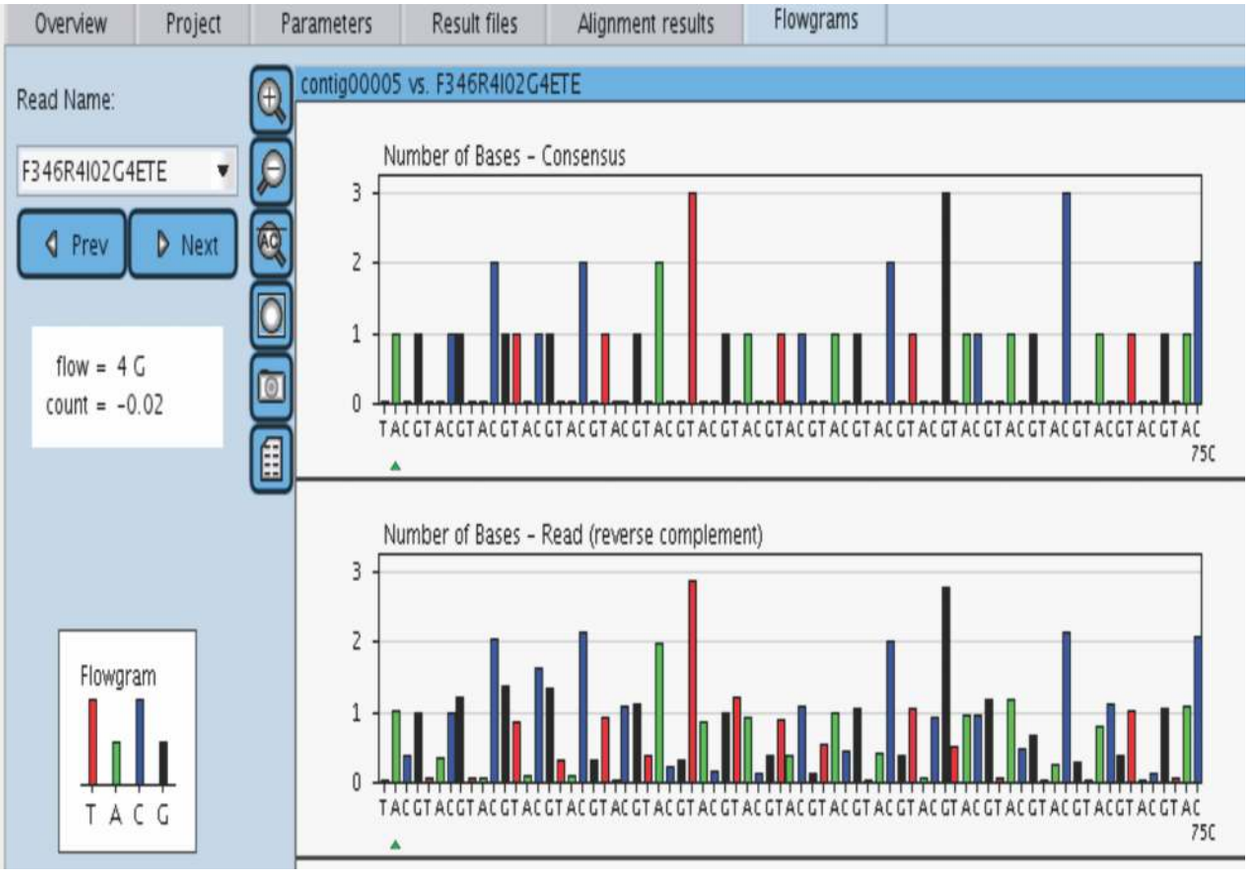


Fig. 7 Consensus contigs generated and its similarity and differences with the individual sequence reads after *de novo* Assembly of 454 sequencing by De novo Assembler. 1, 2, and 3 represents the number of bases and their repeats at a particular position

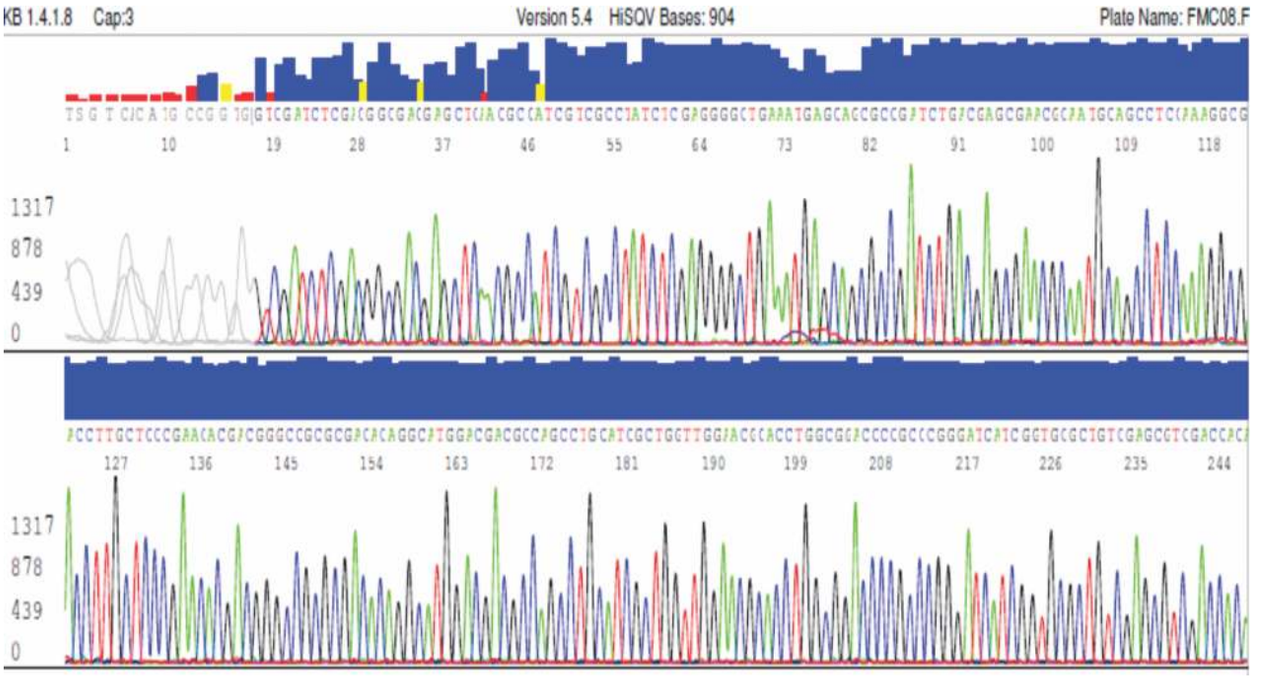


Fig. 8 Quality Electropherogram file of large insert library Sanger Sequencing data

Annotation and Analysis

After the annotation of genome sequences 6742 genes in 4116 transcriptional units were predicted and searched for the functions they perform in *Mesorhizobium ciceri* Ca181 (Fig. 9). Most of genes belong to known functional categories and about 1934 genes are involved in physiological functions, 184 genes in stress management, 1268 genes in transportation, 179 genes in DNA synthesis and repair and 1532 genes were for hypothetical proteins (Fig. 11). Total 34 genes are involved in nitrogen fixation but functions of about 601 genes are still unknown and may be unique to *M. ciceri* Ca 181 (Fig. 12).

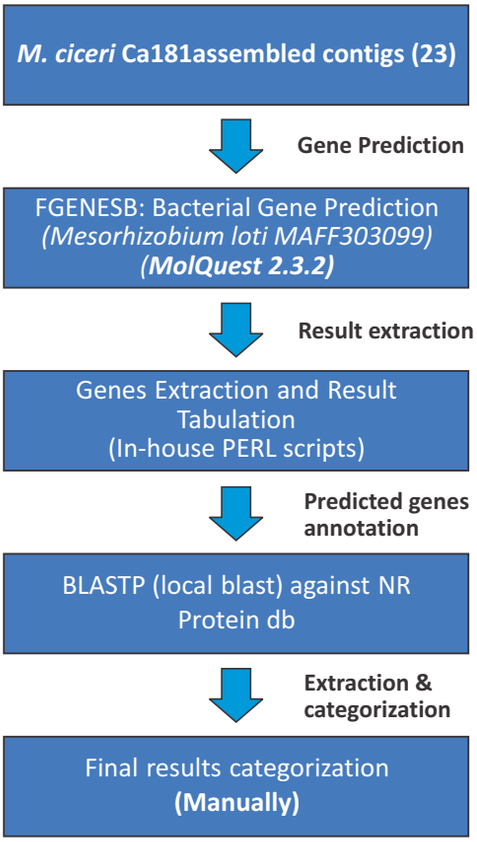


Fig .9 Pipeline for gene prediction and annotation in the genome

S.No.	Scaffold Name	Size	Predicted Genes	Transcription Units
1	Scaffold -01	4160774	4284	2615
2	Scaffold -02	1154896	1166	36
3	Scaffold -03	952708	944	654
4	Scaffold -04	199993	202	18
5	Scaffold -05	43706	46	29
6	Scaffold -06	41138	45	73
7	Scaffold -07	29918	28	124
8	Scaffold -08	16195	15	567
9	Scaffold -09	12846	12	2586
	Total Size		**6742**	**4116**

Table. 6 Scaffolds and their size in base-pairs and number of predicted genes per scaffold

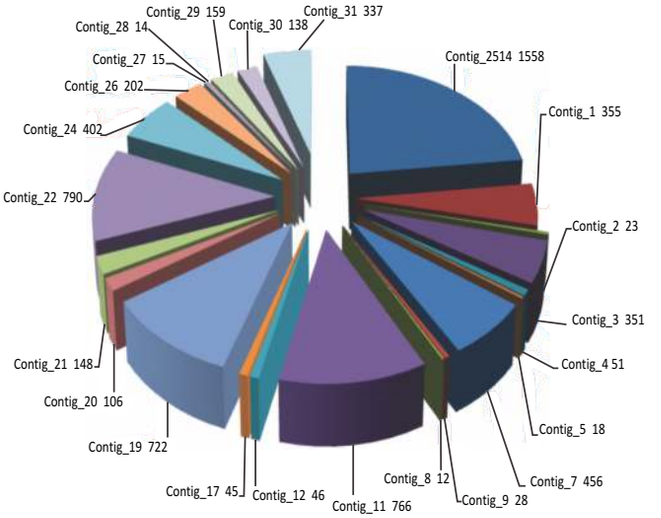


Fig. 10 Distribution of Genes in the Genome of *M. ciceri* Ca181

Type	No. of Genes
Physiological Traits	1934
Stress Response	184
Transportation	1268
Protein Synthesis	523
Nitrogen Fixation	34
Hypothetical	1532
DNA Synthesis & Repair	179
Growth & Development	21
Unclassified	465
Unannotated	602
TOTAL	6742

Table 7: Annotation categories of genes predicted in the *M. ciceri* Ca181 Genome

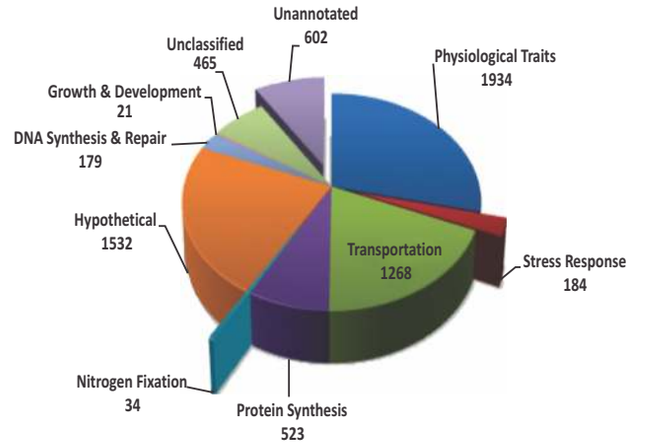
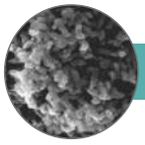


Fig. 11 Genes involved in different functions in the Genome of *M. ciceri* Ca181



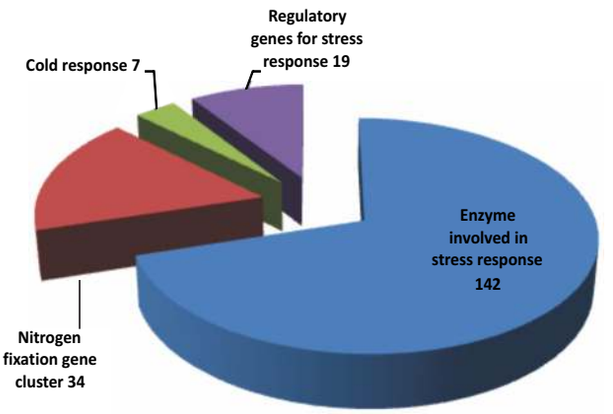


Fig. 12 Important functional clusters of genes of the *M. Ciceri* Ca181 Genome

1	Total Contigs used for gene prediction	**23**
2	Total no of gaps	***22***
3	Total no of gap filled	***16***
4	Total no of scaffold	***9***
5	Total no of gap remaining	***7***

Table 8. *S*ummary details of Assembly of Genome of *M. ciceri* Ca181

Comparative Genome Analysis

The genome sequence of *M. ciceri* Ca181 was compared with a recently published sequenced *M. ciceri* biover biserrulae and earlier published *M. loti* sequence for similarity search, which exhibit only 32% similarity with *M. loti* , while 51% similarity was recorded with *M. ciceri* biserrulae.

Conclusion

The genome database of *M. ciceri* Ca181 will be useful for comparative analysis with other bacterial genomes, and discovery of genes responsible for agronomically useful traits such as competitiveness, stress tolerance, nitrogen fixing ability and growth promoting activity etc. The ongoing AMAAS genomics project will now focus on the gene discovery (functional genomics) aspects and will help Indian Agriculture bio-fertilizer programmes in the following ways:

- Improving the ability of Rhizobium strain to fix nitrogen and thereby resulting in decrease use of chemical fertilizers.
- Enabling the strain to thrive and fix nitrogen under adverse conditions such as saline, alkaline and acidic soils as well as drought and water logging conditions.
- Improving the ability of the strain for better adaptability in soil to outcompete native soil microbial population.
- Improving soil fertility by improving its ability to neutralize the nitrate toxicity, phosphate solubilization and other plant growth promoting activities besides fixing nitrogen in the root nodules.
- Help to make a good microbial bio-fertilizer to enhance crop productivity.

Research Groups :

National Bureau of Agriculturally Important Microorganisms, Mau

D. K. Arora, Director, NBAIM, Mau
A. K. Saxena, Principal Scientist (8/2006-5/2009)
Rajeev Kaushik, Senior Scientist (8/2006-6/2010)
A. K. Srivastava, Senior Scientist
Ram Nageena Singh, RA
Preeti Shahi, SRF
Anu Sharma, SRF
Raghvendra P. Singh, SRF
Avanika Singh, SRF
Dharmesh Singh, SRF (2006-2007)
Rinku Singh ,SRF (2006-2007)

National Research Centre for DNA Fingerprinting, New Delhi

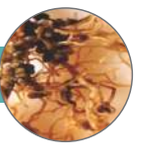
K. V. Bhat, Principal Scientist
Rakesh Singh, Senior Scientist
Ambika Gaiakwaad, Scientist
Ratna Kumari, SRF
Deepali Sharma, SRF

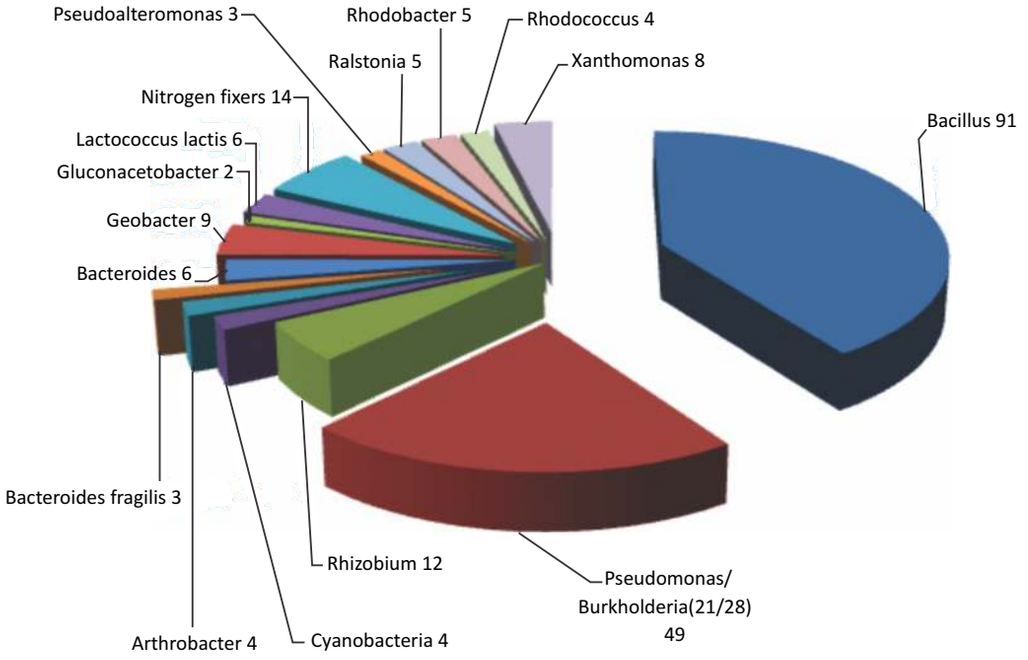
Indian Institute of Vegetable Research, Varanasi

Major Singh, Principal Scientist
Avinash Rai, SRF
Akhilesh Kumar, SRF

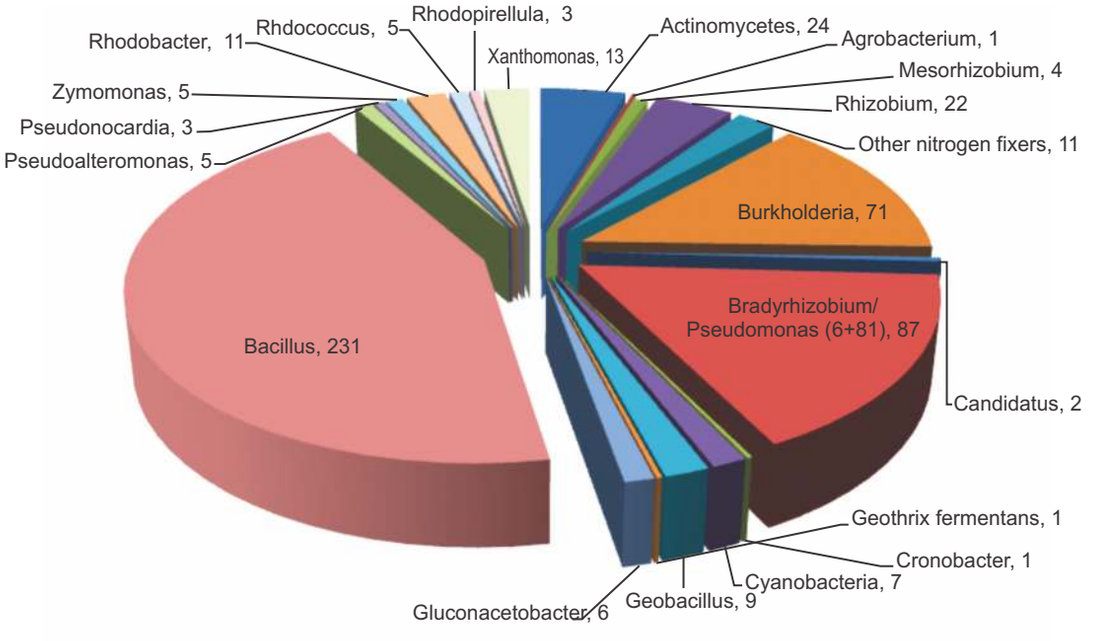
National Research Centre for Plant Biotechnology, New Delhi

N. K. Singh, National Professor
T.R. Sharma, Principal Scientist
Jasdeep Padaria, Senior Scientist
L.R. Kashyap, Principal Scientist (2006 to June 2008)
Aqbal Singh , Principal Scientist (2006 to May 2009)
Sarvjeet Kaur, Principal Scientist
Kanika, Scientist
Ritu Rai, Scientist (2006 to Jan 2008)
Tripti Dogra, SRF
Sujata Chauhan, S.R.F. (2006 to Feb 2009)
Anupama Priyadarshini, SRF
Ajay K. Mahtoo, RA
Deepak K. Gupta, RA
Vivek Dogra, SRF





Complete genome sequenced of Agriculturally Important Microorganisms



Ongoing genome sequencing projects of Agriculturally Important Microorganisms

Compiled by
Dilip K Arora
N. K. Singh
A. K. Srivastava
Ram Nageena Singh

Corresponding Address
Prof. Dilip K. Arora, Director
NATIONAL BUREAU OF AGRICULTURALLY IMPORTANT MICROORGANISMS
Kusmaur, Mau Nath Bhanjan - 275 101 (U.P.)
Phone: 0547-2530080 Fax : 0547-2530381
E.mail : nbaimicar@gmail.com
Visit us at: www.nbaim.org.in

Published by : Prof. Dilip K. Arora, Director, NBAIM, Mau