User Guide: Moringa Genome Database (MoringaGDB)

MoringaGDB includes ten different tabs: Home, Genes, Markers, Blast Tool, LncRNA, CircRNA, miRNA, Userguide, Downloads and Team (**Figure 1**).

1. **Home:** The main page is strategically designed to offer visitors with a detailed presentation of MoringaGDB, clarifying both its nature and significance. This facilitates deeper understanding of fundamental objectives of the database, its range, and the extensive array of information.

Essentially, the homepage not only presents MoringaGDB but also explains the reasons behind its significance to researchers, scientists, and stakeholders in agriculture. It emphasizes its crucial contribution to the progress of moringa crop research and cultivation.

Moringa Database

The Drumstick plant, or Moringa oleifera, truly is a marvel of nature, thriving in harsh conditions while offering a plethora of essential nutrients and medicinal benefits. Its high protein content, low levels of fat and fiber, and abundance of bioactive compounds make it a valuable resource in both nutrition and pharmaceutical industries.
The various therapeutic properties found in Moringa, such as anti-inflammatory and antimicrobial effects, have sparked growing interest in its potential to combat a wide range of health issues.Our study on transcriptome analysis of the Moringa oleifera cultivar 'Bhagya' is a significant contribution to understanding the genetic mechanisms underlying its remarkable traits. The creation of the Moringa Genome Database (MoringaGDB) is a commendable effort to consolidate and disseminate valuable genomic information to breeders and researchers.

With its comprehensive insights into secondary metabolites, genes, and genetic markers, MoringaGDB has the potential to significantly advance research and enhance the productivity of Moringa cultivation.

This online platform provides extensive insights into medicinally significant secondary metabolites/genes and long non-coding RNA, including their expression levels, circular RNA, sequence information, functions, pathways, and genetic markers (such as SSRs, SNPs, and InDels), along with their chromosomal locations.

The three-tier architecture of MoringaGDB, utilizing HTML/Java for the user interface, PHP for server-side scripting and database connectivity, and MySQL as the backend repository, ensures efficient access and management of genomic data. This database holds promise for facilitating further studies on Moringa and accelerating its utilization for various applications, from agriculture to medicine.

Funded by

This Web Resource is developed under the project "Establishment of Centre for Bioinformatics and Computational Biology in Agriculture-BIC" funded by the Department of Biotechnology, Govt. of India, New Delhi, India

Contact Us on Social Media
Facebook
Twitter

Address
ICAR-Indian Agricultural Statistics Research Institute,
Library Avenue, Pusa, New Delhi-110012

Figure 1. The home page of Moringa Genome Database (MoringaGDB)

2**. Genes:** It includes s dropdown menu which consists of DEGs and pathways from different tissues of Moringa. The Page consists of two options, where user can select the tissue and regulation of genes related to their expression (**Figure 2**)

Results for Flower_leaf and Up-regulated

Gene_id	Sequence_Name	FPKM	FPKM_2	log2_fold_change	P_value	Blast_hit	Go	Enzyme	Interpro_ID
XLOC_000205	JAJFZO010000008.1_2979-6562	0.125595	1.17404	3.22463	0.0021	BLAST	View	View	-
XLOC_000219	JAJFZO010000009.1_21083-22145	0.543253	54.5984	6.65109	0.0001	BLAST	View	View	IPR000484
XLOC_000224	JAJFZO010000009.1_37491-40370	0.115323	0.932429	3.01531	0.0102	BLAST	View	View	IPR007081
XLOC_000231	JAJFZO010000011.1_3736-8858	0.134506	0.763006	2.50402	0.00065	BLAST	View	View	-

Sequence for: JAJFZO010000008.1_2979-6562

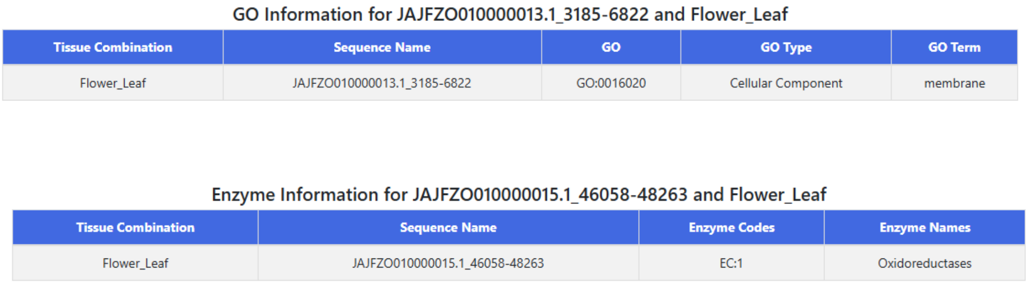
```
atgctacgcgagggccaaatcgttgggggatttcgagcatgcggagtgcttcgtcgcaggccctcgaggcggcggacgcggtgctggatcccgacgcccaggcttgc
ctaggcggactcgttgcgcagcgcgatcgacgacacgatcgtcggccgttagcaatgagcgagccacacggccagcctgggcaaaggcgtgacacccctctgccaggc
ctcatcggtgggaaaacgtgggcaacgaaacgagggaaa
```

BLAST Hit Details for Sequence: JAJFZO010000008.1_2979-6562 and Tissue Combination: Flower_Leaf

Tissue Combination	Sequence Name	Description	Sequence Length	Blast Hits Count	Blast Min E-Value	Blast Similarity Mean
Flower_Leaf	JAJFZO010000008.1_2979-6562	hypothetical protein DVH24_023838 [Malus domestica]	3583	20	8.40E-70	73.45

Figure 2: Showing the Differentially Expressed Genes (DEGs) and their biological functions. Users have the option to search for information by selecting the tissue condition or using Gene_ID. By clicking on the Gene_ID, users can retrieve the sequence in fasta format. The blast button provides information about the function of the scaffold along with sequence similarity.

This page provides details on genes that exhibit differential expression (upregulated or downregulated). The expression levels of these genes are presented in the form of FPKM (fragments per kilobase of transcript per million mapped reads) values. Furthermore, it provides functional annotation details such as blast results, domains, enzyme class, enzyme commission (EC) number, and information on Gene Ontology (GO) categories encompassing biological processes, molecular functions, and cellular components (**Figure 3**). The second drop down in genes page provides information on pathway expressed in different tissues (**Figure 4**).



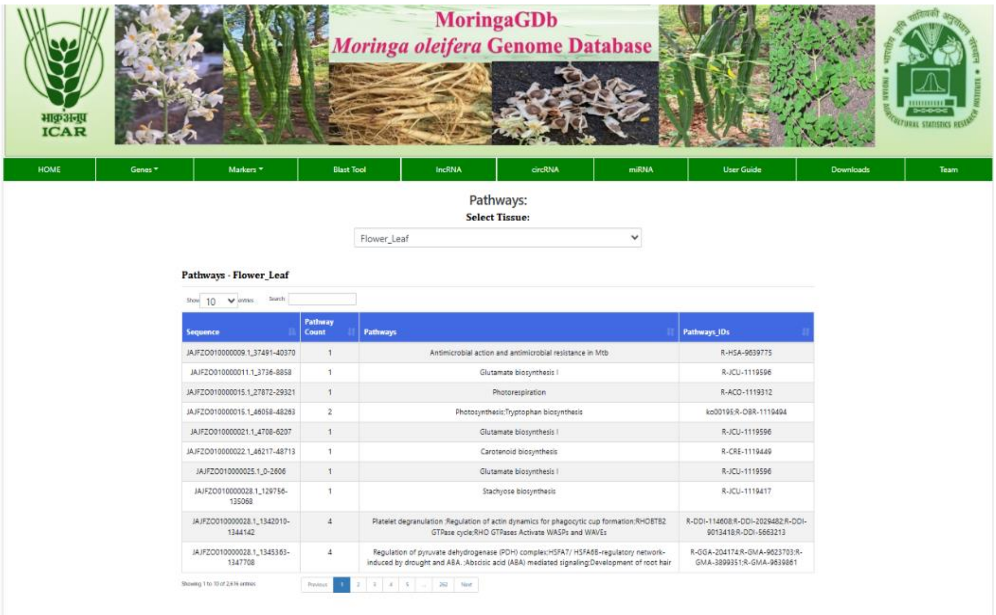
GO Information for JAJFZO010000013.1_3185-6822 and Flower_Leaf

Tissue Combination	Sequence Name	GO	GO Type	GO Term
Flower_Leaf	JAJFZO010000013.1_3185-6822	GO:0016020	Cellular Component	membrane

Enzyme Information for JAJFZO010000015.1_46058-48263 and Flower_Leaf

Tissue Combination	Sequence Name	Enzyme Codes	Enzyme Names
Flower_Leaf	JAJFZO010000015.1_46058-48263	EC:1	Oxidoreductases

Figure 3: Information about gene ontology and enzyme information foe gene scaffold is provided in the view tab present on the main page.



Pathways:

Select Tissue:

Flower_Leaf

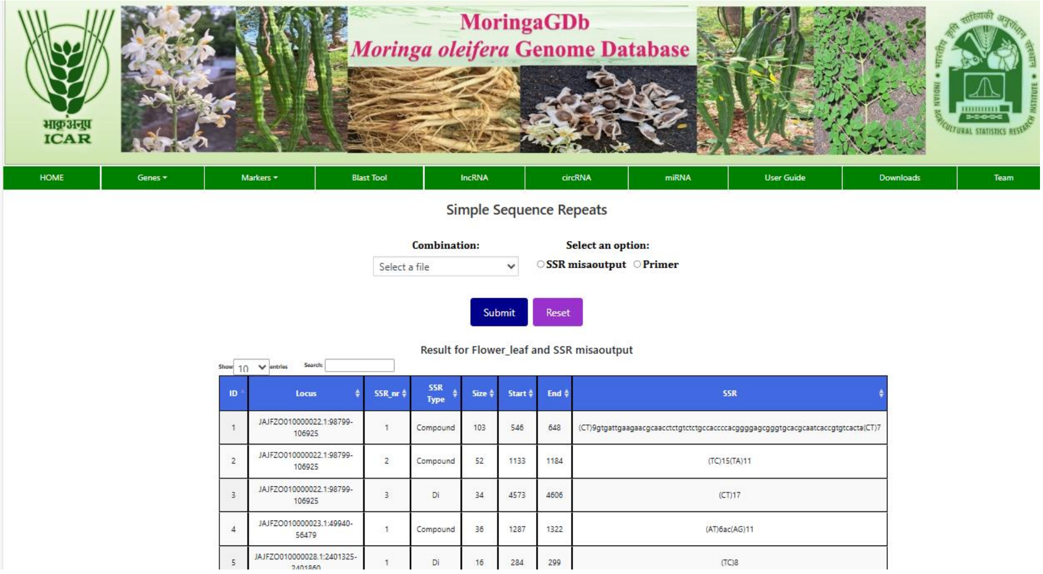
Pathways - Flower_Leaf

Sequence	Pathway Count	Pathways	Pathways_IDs
JAJFZO010000009.1_37491-40370	1	Antimicrobial action and antimicrobial resistance in Mtb	R-HSA-9639775
JAJFZO010000011.1_3736-8858	1	Glutamate biosynthesis I	R-JCU-1119596
JAJFZO010000015.1_27872-29321	1	Photorespiration	R-ACO-1119312
JAJFZO010000015.1_46058-48263	2	Photosynthesis;Tryptophan biosynthesis	ko00195;R-OBR-1119494
JAJFZO010000021.1_4708-6207	1	Glutamate biosynthesis I	R-JCU-1119596
JAJFZO010000022.1_46217-48713	1	Carotenoid biosynthesis	R-CRE-1119449
JAJFZO010000025.1_0-2606	1	Glutamate biosynthesis I	R-JCU-1119596
JAJFZO010000028.1_129756-135068	1	Stachyose biosynthesis	R-JCU-1119417
JAJFZO010000028.1_1342010-1344142	4	Platelet degranulation ;Regulation of actin dynamics for phagocytic cup formation;RHOBTB2 GTPase cycle;RHO GTPases Activate WASPs and WAVEs	R-DDI-114608;R-DDI-2029482;R-DDI-9013418;R-DDI-5663213
JAJFZO010000028.1_1345363-1347708	4	Regulation of pyruvate dehydrogenase (PDH) complex;HSFA7/ HSFA6B-regulatory network-induced by drought and ABA. ;Abscisic acid (ABA) mediated signaling;Development of root hair	R-GGA-204174;R-GMA-9623703;R-GMA-3899351;R-GMA-9639861

Showing 1 to 10 of 2,614 entries

Figure 4: Showing the diverse pathways associated with each gene. This tab ID ‘s from Reactome and KEGG pathways associated with each gene.

3. **Markers**: It includes a dropdown menu consisting of Simple Sequence Repeats (SSR), Single Nucleotide Polymorphism (SNP) and Insertions and Deletions (InDels). The SSR page include two sets of primers and a gene sequence (**Figure 5, 6 and 7**)



Simple Sequence Repeats

Combination:

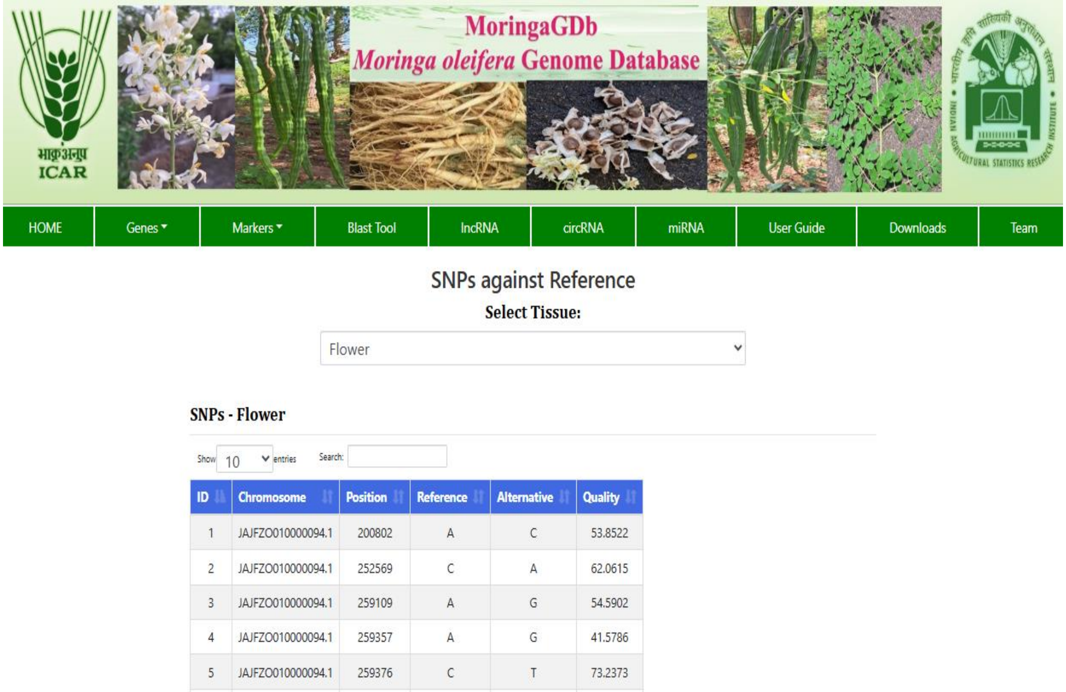
Select an option:

Result for Flower_leaf and SSR misaoutput

ID	Locus	SSR_nr	SSR Type	Size	Start	End	SSR
1	JAJFZO010000022.1:98799-106925	1	Compound	103	546	648	(CT)9gtgattgaagaacgcaacctctgtctctgccaccccacggggagcgggtgcacgcaatcaccgtgtcacta(CT)7
2	JAJFZO010000022.1:98799-106925	2	Compound	52	1133	1184	(TC)15(TA)11
3	JAJFZO010000022.1:98799-106925	3	Di	34	4573	4606	(CT)17
4	JAJFZO010000023.1:49940-56479	1	Compound	36	1287	1322	(AT)6ac(AG)11
5	JAJFZO010000028.1:2401325-2401860	1	Di	16	284	299	(TC)8

Figure 5: Displaying the simple sequence repeats (SSRs). Locus allows viewing the genomic location of SSR using the genome browser and also has the option to generate primers by clicking on "Primers".

In addition, it provides information on genic single nucleotide polymorphisms (SNPs) and InDel markers, along with their genomic positions determined through mapping with a reference genome.



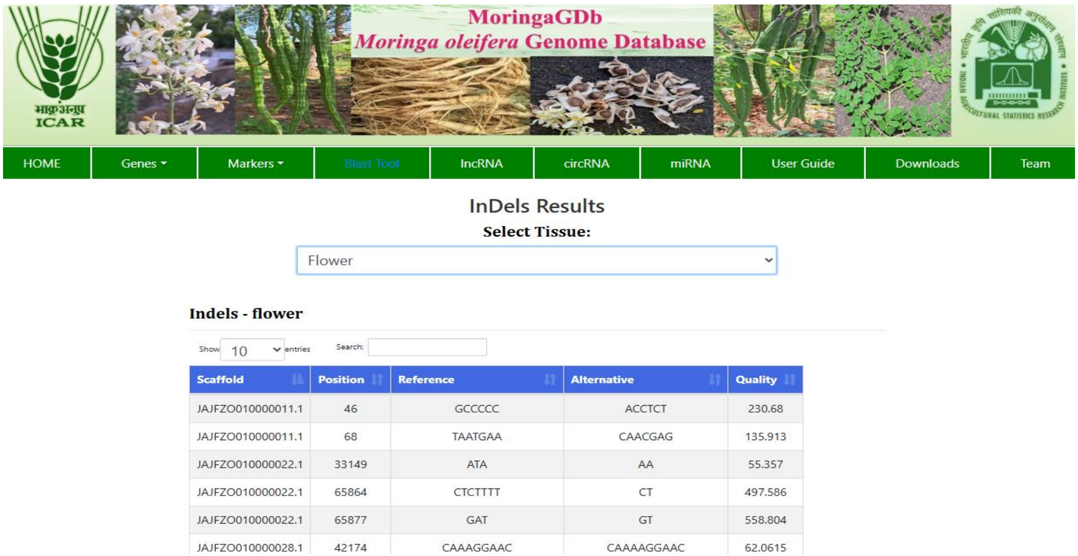
SNPs against Reference

Select Tissue: Flower

SNPs - Flower

ID	Chromosome	Position	Reference	Alternative	Quality
1	JAJFZO010000094.1	200802	A	C	53.8522
2	JAJFZO010000094.1	252569	C	A	62.0615
3	JAJFZO010000094.1	259109	A	G	54.5902
4	JAJFZO010000094.1	259357	A	G	41.5786
5	JAJFZO010000094.1	259376	C	T	73.2373

Figure 6. Showing the single nucleotide polymorphisms (SNPs) against a reference genome. Locus allows users to view the genomic location of SNP present in different moringa tissues in the genome browser.



InDels Results

Select Tissue: Flower

Indels - flower

Show 10 entries Search:

Scaffold	Position	Reference	Alternative	Quality
JAJFZO010000011.1	46	GCCCCC	ACCTCT	230.68
JAJFZO010000011.1	68	TAATGAA	CAACGAG	135.913
JAJFZO010000022.1	33149	ATA	AA	55.357
JAJFZO010000022.1	65864	CTCTTTT	CT	497.586
JAJFZO010000022.1	65877	GAT	GT	558.804
JAJFZO010000028.1	42174	CAAAGGAAC	CAAAAGGAAC	62.0615

Figure 7. Showing the insertions and deletions (InDels) against a reference genome. Genome browser allows users to view the genomic location of an InDel.

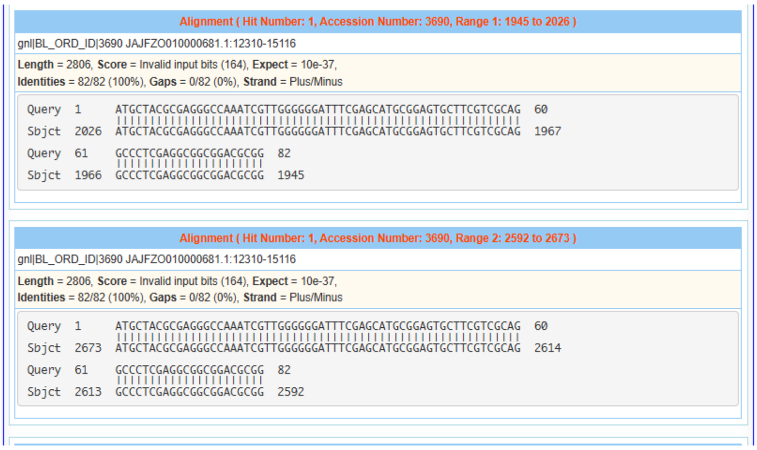
4. **Blast Tool-** Users can conduct a BLAST search to determine the similarity of their sequence with the moringa genome. They can also access information about various moringa tissues using the gene ID from our database. The advanced search has various parameters such as maximum target sequences, word size (6-256), different matrix type and different filtering options. The advance option displays the sequence alignment with query sequence so that users can explore the query coverage.

BLAST - local BLAST

Program	blastn
Version	BLASTN 2.15.0+
Reference	Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
Database	db/flower_leaf_DEG
Query ID	Query_1
Definition	JAJFZO010000008.1:2979-6562
Length	92
E-value	10
Gap Open	0
Gap Extend	0
Filter	L;m;

Sequences producing significant alignments:

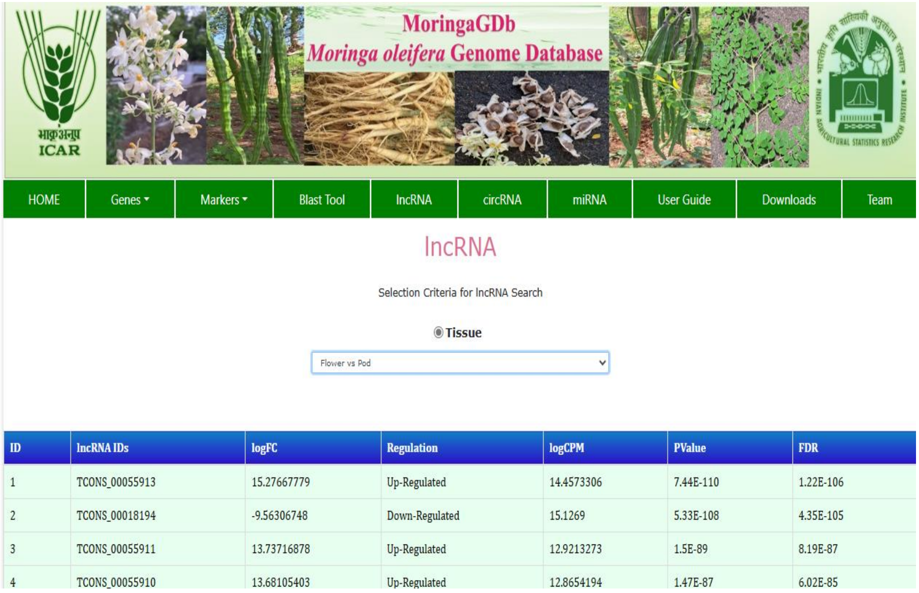
Sequence Description	Query cover	E Value	Ident	Accession
JAJFZO010000681.1:12310-15116	89%	10e-38	100%	3690
JAJFZO010000633.1:876-4157	89%	10e-38	100%	3481
JAJFZO010000559.1:193-4474	89%	10e-38	100%	2631
JAJFZO010000475.1:3588-7525	89%	10e-38	100%	2571
JAJFZO010000464.1:44-3507	89%	10e-38	100%	2545
JAJFZO010000438.1:0-5338	89%	10e-38	100%	2505



```
Alignment ( Hit Number: 1, Accession Number: 3690, Range 1: 1945 to 2026 )
gnl|BL_ORD_ID|3690 JAJFZO010000681.1:12310-15116
Length = 2806, Score = Invalid input bits (164), Expect = 10e-37,
Identities = 82/82 (100%), Gaps = 0/82 (0%), Strand = Plus/Minus

Query  1     ATGCTACGCGAGGGCCAAATCGTTGGGGGGATTTCGAGCATGCGGAGTGCTTCGTCGCAG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2026  ATGCTACGCGAGGGCCAAATCGTTGGGGGGATTTCGAGCATGCGGAGTGCTTCGTCGCAG  1967

Query  61    GCCCTCGAGGCGGCGGACGCGG  82
             ||||||||||||||||||||||
Sbjct  1966  GCCCTCGAGGCGGCGGACGCGG  1945

Alignment ( Hit Number: 1, Accession Number: 3690, Range 2: 2592 to 2673 )
gnl|BL_ORD_ID|3690 JAJFZO010000681.1:12310-15116
Length = 2806, Score = Invalid input bits (164), Expect = 10e-37,
Identities = 82/82 (100%), Gaps = 0/82 (0%), Strand = Plus/Minus

Query  1     ATGCTACGCGAGGGCCAAATCGTTGGGGGGATTTCGAGCATGCGGAGTGCTTCGTCGCAG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2673  ATGCTACGCGAGGGCCAAATCGTTGGGGGGATTTCGAGCATGCGGAGTGCTTCGTCGCAG  2614

Query  61    GCCCTCGAGGCGGCGGACGCGG  82
             ||||||||||||||||||||||
Sbjct  2613  GCCCTCGAGGCGGCGGACGCGG  2592
```

Figure 7: Various parameters for running Blast tool

5. **LncRNA**: This tab provides information of differentially expressed long non coding RNAs present across different tissues along with their logfold change, type of regulation and p value.



lncRNA

Selection Criteria for lncRNA Search

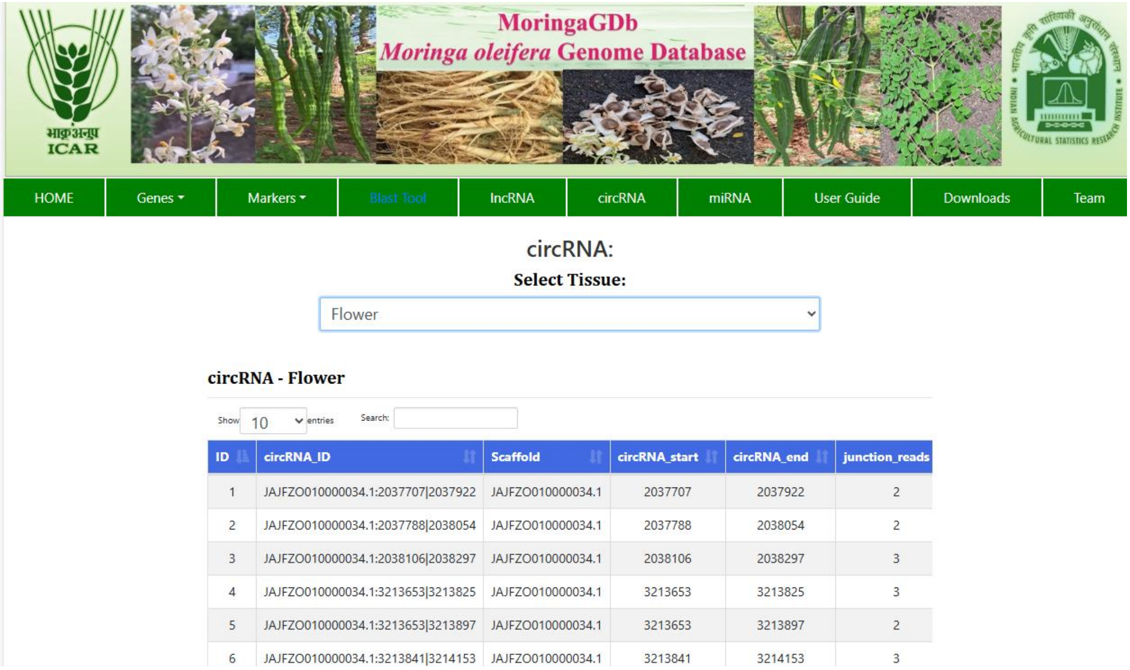
● Tissue

Flower vs Pod

ID	lncRNA IDs	logFC	Regulation	logCPM	PValue	FDR
1	TCONS_00055913	15.27667779	Up-Regulated	14.4573306	7.44E-110	1.22E-106
2	TCONS_00018194	-9.56306748	Down-Regulated	15.1269	5.33E-108	4.35E-105
3	TCONS_00055911	13.73716878	Up-Regulated	12.9213273	1.5E-89	8.19E-87
4	TCONS_00055910	13.68105403	Up-Regulated	12.8654194	1.47E-87	6.02E-85

Figure 8: Showing the differentially present long non coding RNA and their biological functions. Users have the option to search for information by selecting the tissue condition or using lncRNA_ID.

6. **CircRNA:** The page includes circular RNA of different tissues *viz*. Leaf, Root, Stem, Flower and Pod. The page also provides information about different types of circular RNA (Exonic, Intergenic and Intronic) along with their start and end positions (**Figure 9**).



circRNA:

Select Tissue: Flower

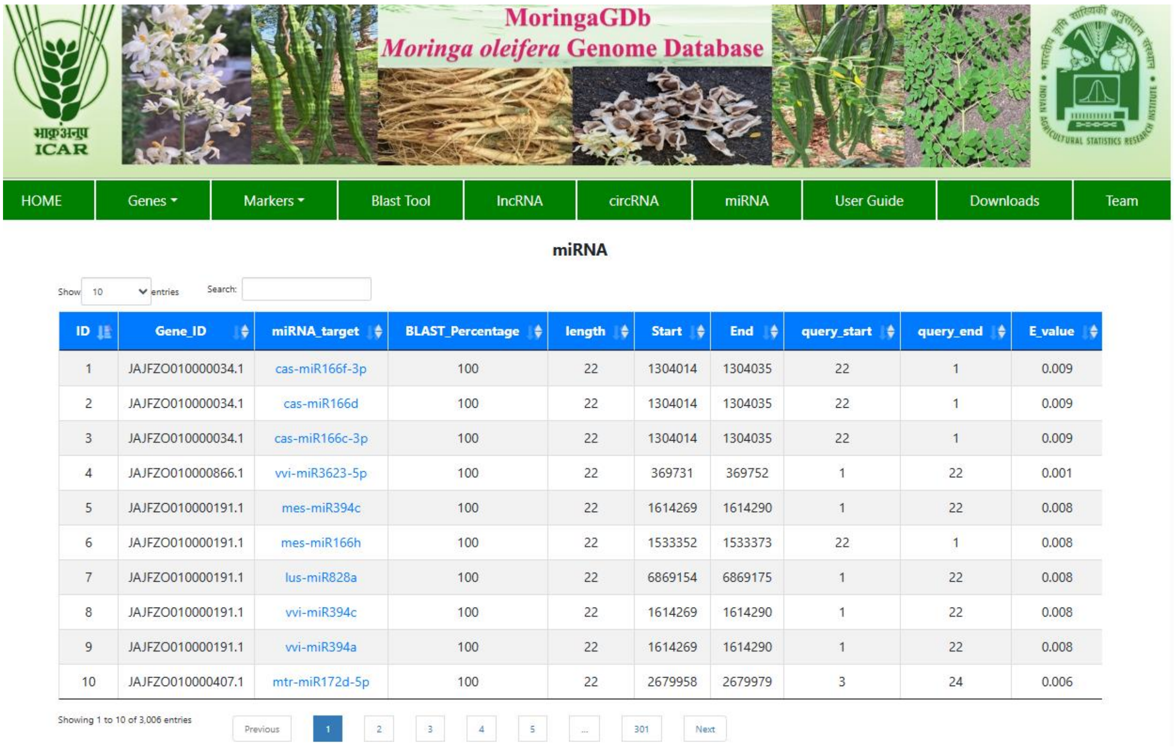
circRNA - Flower

Show 10 entries

ID	circRNA_ID	Scaffold	circRNA_start	circRNA_end	junction_reads
1	JAJFZO010000034.1:2037707|2037922	JAJFZO010000034.1	2037707	2037922	2
2	JAJFZO010000034.1:2037788|2038054	JAJFZO010000034.1	2037788	2038054	2
3	JAJFZO010000034.1:2038106|2038297	JAJFZO010000034.1	2038106	2038297	3
4	JAJFZO010000034.1:3213653|3213825	JAJFZO010000034.1	3213653	3213825	3
5	JAJFZO010000034.1:3213653|3213897	JAJFZO010000034.1	3213653	3213897	2
6	JAJFZO010000034.1:3213841|3214153	JAJFZO010000034.1	3213841	3214153	3

Figure 9: Tissue-wise sample information on circRNA ID's.

7. **miRNA:** This page has Gene ID's and their miRNA target in various plant species with percent similarity along with their e values.



miRNA

Show 10 entries

ID	Gene_ID	miRNA_target	BLAST_Percentage	length	Start	End	query_start	query_end	E_value
1	JAJFZO010000034.1	cas-miR166f-3p	100	22	1304014	1304035	22	1	0.009
2	JAJFZO010000034.1	cas-miR166d	100	22	1304014	1304035	22	1	0.009
3	JAJFZO010000034.1	cas-miR166c-3p	100	22	1304014	1304035	22	1	0.009
4	JAJFZO010000866.1	vvi-miR3623-5p	100	22	369731	369752	1	22	0.001
5	JAJFZO010000191.1	mes-miR394c	100	22	1614269	1614290	1	22	0.008
6	JAJFZO010000191.1	mes-miR166h	100	22	1533352	1533373	22	1	0.008
7	JAJFZO010000191.1	lus-miR828a	100	22	6869154	6869175	1	22	0.008
8	JAJFZO010000191.1	vvi-miR394c	100	22	1614269	1614290	1	22	0.008
9	JAJFZO010000191.1	vvi-miR394a	100	22	1614269	1614290	1	22	0.008
10	JAJFZO010000407.1	mtr-miR172d-5p	100	22	2679958	2679979	3	24	0.006

Showing 1 to 10 of 3,006 entries

Figure 10: The miRNA targets from various plant species are accessible through a link that is connected to miRBase.

8. **Downloads:** This tab contains the links to the files which contains information of matched CircRNA Id’s from our dataset with PlantCircBase.

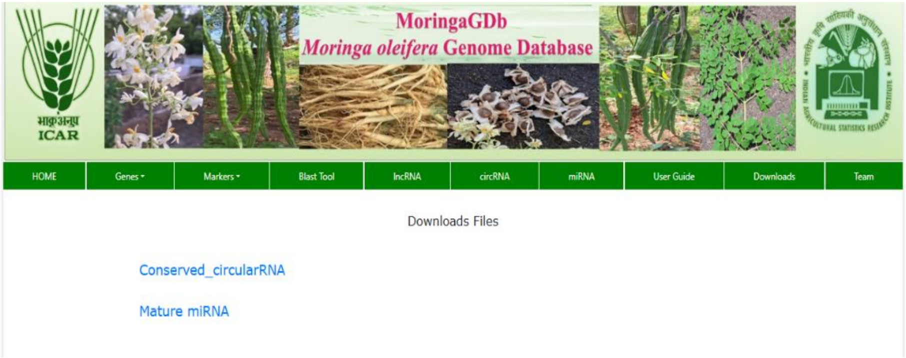


Figure 11: Conservation analysis of identified circRNAs from our dataset. The third column contains circular RNA Ids of different species. It also contains information about mature miRNA’s present in our dataset.