In silico development of CRISPR/CAS 9 Construct for *Oryza sativa* subsp. *indica*

Anis Afuza Md Yusof
International Islamic University of Malaysia

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA
Garden of Knowledge and Virtue

LEADING THE WAY
KHALĪFAH • AMĀNAH • IQRA' • RAḤMATAN LIL-ĀLAMĪN

GREATER GOMBAK

REGIONAL CENTRE OF EXPERTISE ON EDUCATION FOR SUSTAINABLE DEVELOPMENT
ACKNOWLEDGED BY
UNITED NATIONS UNIVERSITY

INTRODUCTION

❑ One of the most important staple foods in the world especially in Asian countries such as Malaysia

❑ Contains 3 subspecies; *O. sativa* subsp. *indica*, *O. sativa* subsp. *japonica* & O. sativa subsp. javanica.

❑ Many studies & researches have been done to this commercially important crop to enable further improvement in order to maximize national rice self-sufficiency level (SSL) & meet high global demand

❑ Diverge nation's aim in achieving food security at the national level (rice self-sufficiency level).

❑ Climate change – limiting factors in rice production

❑ In 2016, The Star reported more than 7,500 ha of paddy fields in Malaysia were affected by drought.

❑ Due to high demand in solving this problem; CRISPR - the gene of interest could be rightly targeted towards improving rice survival ability in drought stress environment.

❑ This study targets the gene of interest (GOI), *OsSCE1* using CRISPR/Cas9 technology.

METHODOLOGY

PART 1

DNA Database

OsSCE1 Prediction

OsSCE1 Annotation

BLAST2GO:OmicsBox:
BLAST
InterProScan
Gene Ontology (GO) mapping
GO Mapping

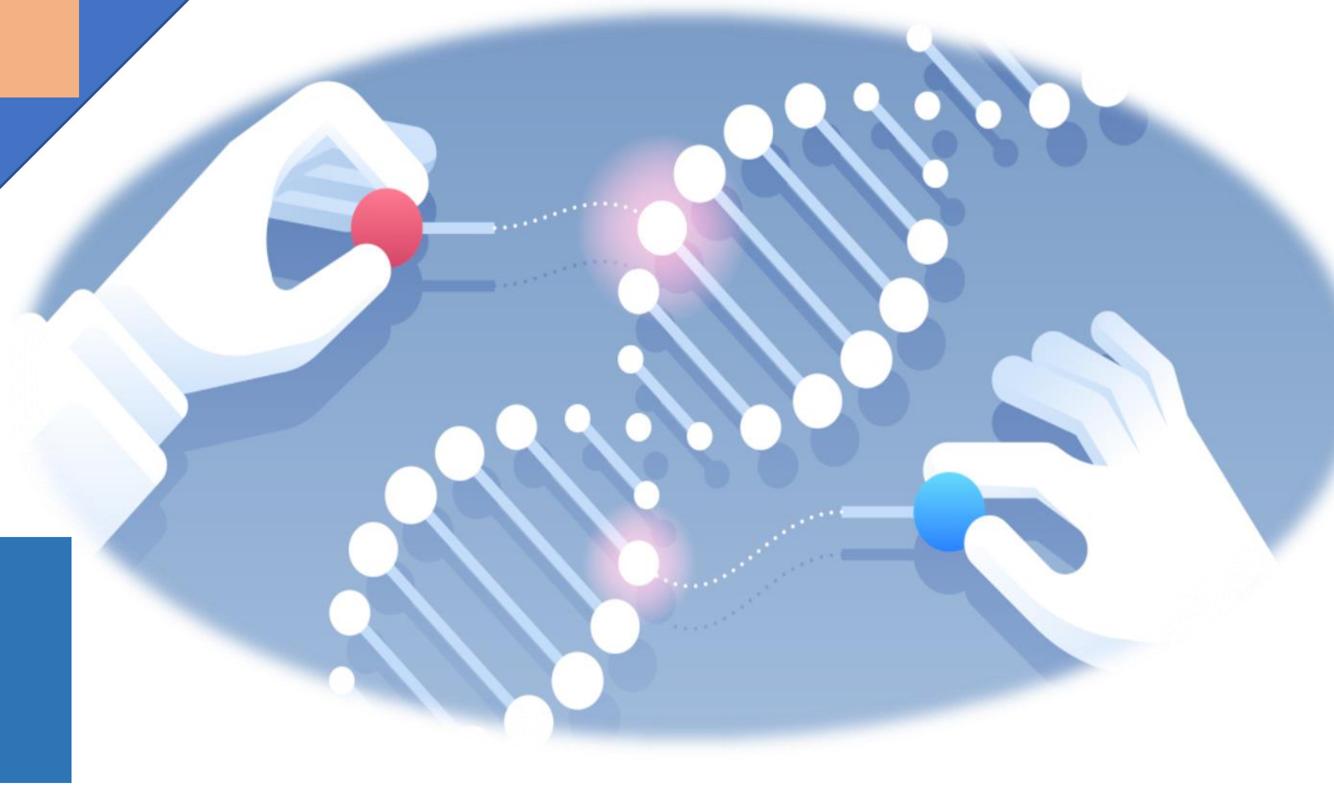
PART 2

DNA extraction using CTAB method

Primer design

PCR

Validate via Sequencing



sgRNA Design

Vector Construction

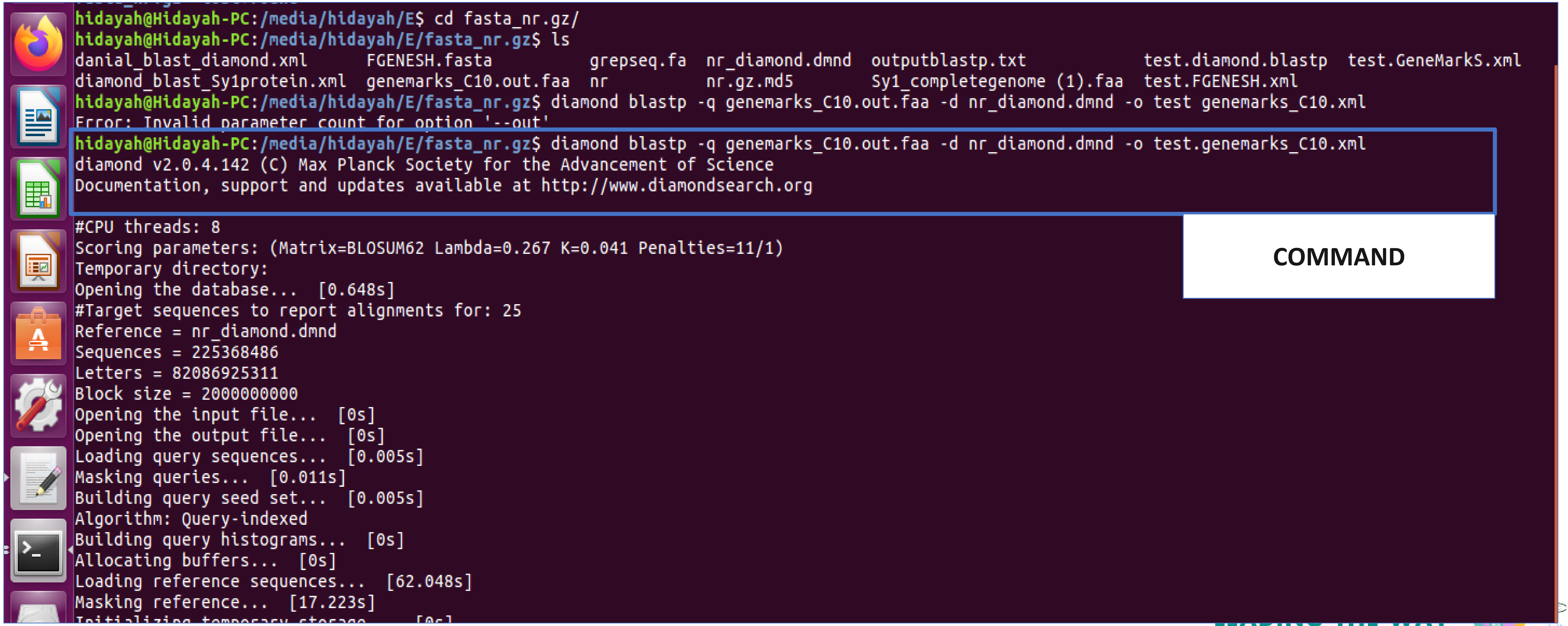
OsSCE1 Gene Prediction

The sequence for *Oryza sativa* ~~subsp. *indica*~~ for chromosome 10 genome was obtained and downloaded from the NCBI database in FASTA format.

Gene prediction was conducted using gene prediction software such as GeneMarkS & FGENESH. The gene prediction was executed by inserting the FASTA sequence obtained earlier from NCBI.

Predicted gene sequences were downloaded in FASTA format from the respective webpages and imported into LINUX terminal and homology search was done using BLAST DIAMOND.

The sequences was imported into **LINUX (UBUNTU)** & blasted using DIAMOND BLAST. The output was set in **XML format** (easy to import in Blast2GO).



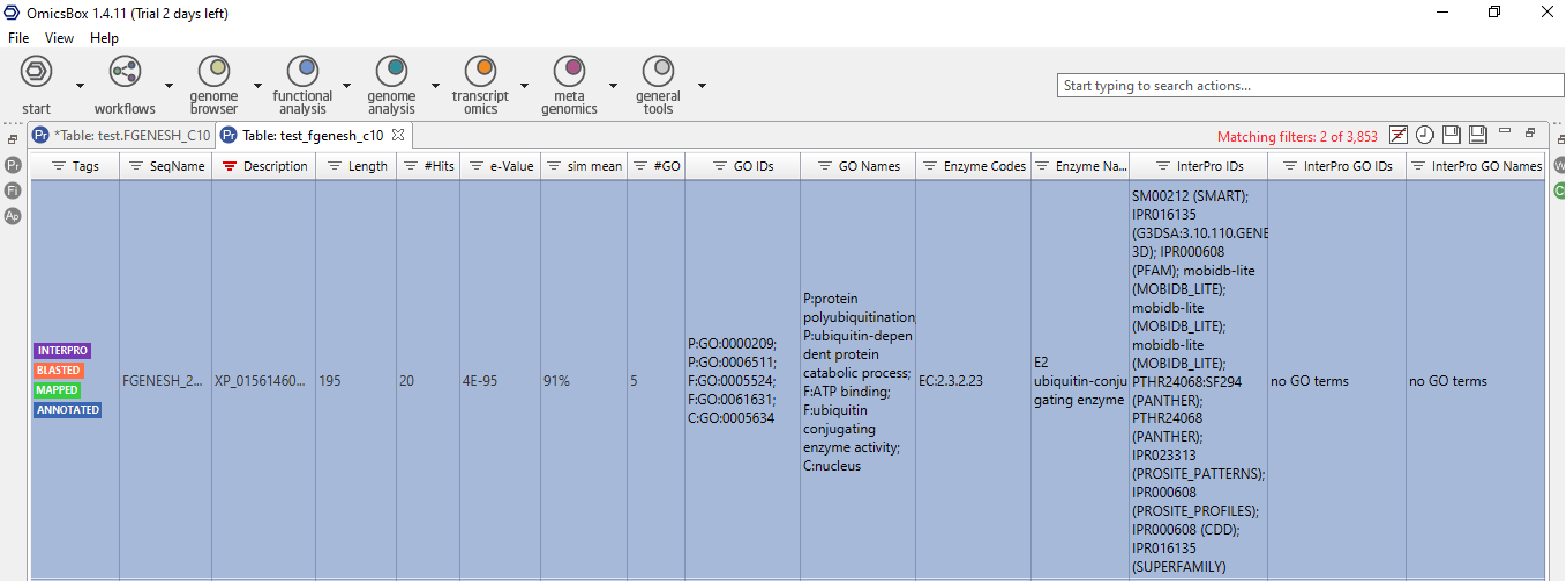
```
hidayah@Hidayah-PC:/media/hidayah/E$ cd fasta_nr.gz/
hidayah@Hidayah-PC:/media/hidayah/E/fasta_nr.gz$ ls
danial_blast_diamond.xml      FGENESH.fasta          grepseq.fa  nr_diamond.dmnd  outputblastp.txt            test.diamond.blastp  test.GeneMarkS.xml
diamond_blast_Sy1protein.xml  genemarks_C10.out.faa  nr          nr.gz.md5        Sy1_completegenome (1).faa  test.FGENESH.xml
hidayah@Hidayah-PC:/media/hidayah/E/fasta_nr.gz$ diamond blastp -q genemarks_C10.out.faa -d nr_diamond.dmnd -o test genemarks_C10.xml
Error: Invalid parameter count for option '--out'
hidayah@Hidayah-PC:/media/hidayah/E/fasta_nr.gz$ diamond blastp -q genemarks_C10.out.faa -d nr_diamond.dmnd -o test.genemarks_C10.xml
diamond v2.0.4.142 (C) Max Planck Society for the Advancement of Science
Documentation, support and updates available at http://www.diamondsearch.org

#CPU threads: 8
Scoring parameters: (Matrix=BLOSUM62 Lambda=0.267 K=0.041 Penalties=11/1)
Temporary directory:
Opening the database...  [0.648s]
#Target sequences to report alignments for: 25
Reference = nr_diamond.dmnd
Sequences = 225368486
Letters = 82086925311
Block size = 2000000000
Opening the input file...  [0s]
Opening the output file...  [0s]
Loading query sequences...  [0.005s]
Masking queries...  [0.011s]
Building query seed set...  [0.005s]
Algorithm: Query-indexed
Building query histograms...  [0s]
Allocating buffers...  [0s]
Loading reference sequences...  [62.048s]
Masking reference...  [17.223s]
Initializing temporary storage...  [0s]
```

COMMAND

OsSCE1 Gene Annotation

Predicted gene sequences were annotated using BLAST2GO: OmicsBox (version 1.4.11) software.



Tags	SeqName	Description	Length	#Hits	e-Value	sim mean	#GO	GO IDs	GO Names	Enzyme Codes	Enzyme Na...	InterPro IDs	InterPro GO IDs	InterPro GO Names
INTERPRO BLASTED MAPPED ANNOTATED	FGENESH_2...	XP_01561460...	195	20	4E-95	91%	5	P:GO:0000209; P:GO:0006511; F:GO:0005524; F:GO:0061631; C:GO:0005634	P:protein polyubiquitination; P:ubiquitin-dependent protein catabolic process; F:ATP binding; F:ubiquitin conjugating enzyme activity; C:nucleus	EC:2.3.2.23	E2 ubiquitin-conjugating enzyme	SM00212 (SMART); IPR016135 (G3DSA:3.10.110.GENE3D); IPR000608 (PFAM); mobidb-lite (MOBIDB_LITE); mobidb-lite (MOBIDB_LITE); mobidb-lite (MOBIDB_LITE); PTHR24068:SF294 (PANTHER); PTHR24068 (PANTHER); IPR023313 (PROSITE_PATTERNS); IPR000608 (PROSITE_PROFILES); IPR000608 (CDD); IPR016135 (SUPERFAMILY)	no GO terms	no GO terms

Validation via Wet Lab

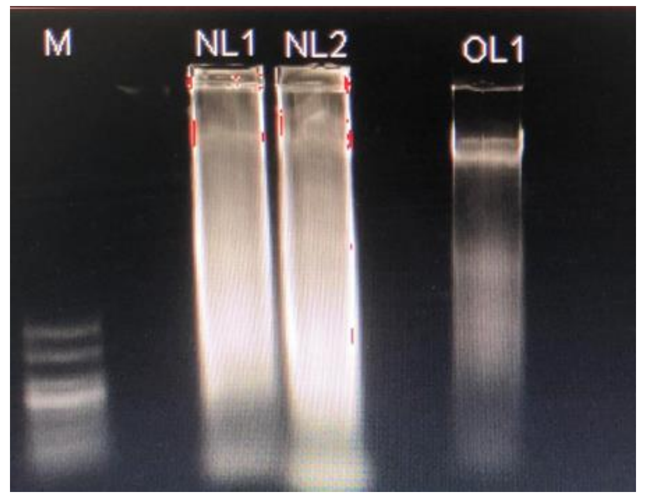
PART 2

DNA Extraction using CTAB Method

Primer Design

PCR

Sequencing



If possible pls put some picture
- your prime seq
- gel of your PCR
- your sequencing

```
>NL1.RVRS

GTAAGGAACTGAATGTTGTAAGATTTAGTTGCTGAAAGATGGATTATAGTCACCATCACTACTGT
CTTCATAATCAACATCACTTCCAAAACACTTTATCTTGCATATGGGGGATATATAGTAATCTTC
AGTGGCGGACCCAGGATTTCTTCATGCCCTAGGCCAAATCTAATTGCTACTATCCCATACCATAA
ATGAGACCCACACTAATACATAATAAGTACAATACAAATTATTATAGTTTCGAAAGGAAAAAAAC
AACAACGCAAGAAACAGCATCTAACCTCAAAATGACAGCTATTTGCTCTTTGAAAGATGCATCTC
CATCATACTCATCAACAAAAATGGCGGATTTTTTTTCCACGGAAAGCATTTAGAATGGTTTTACC
CGTCTCAATTGCACAAGATTTACACTCTAGATTTTATATTTTTAGTGAAGTGAATAGATGATTCC
CGGGAGAATTTTGAAATGTTTCTGGTCATCTTTATTCTTCTCTTTATACCCCATCAATCATCTAT
AGATTTCATCGCTAAGCCTCTTCACCTTCTATCTCTATTAAGTTTAAAACAAATTGAATGGAATC
AGAAGATCTATAGGTGATCTAGGCCATTTAGGCATTTGAGAAATGGGAATACCAGAAATAGATTA
TGAAAAACTCACCTGTCAAGGTCCTCGTCCGCTCCTTGCCCGACAGAGTGGAGGTCTGAGACTCT
GAGCACCTGCCGCTTGCCGCAGGCGCACCGTGCAGGGGGGTTGTGCAGAAAGAGCCTTGTTGCCG
GGCCGCCGAGCCAGCGAGAGGAGCTGAGGAGGCAGTGGACATGGACGGAGGGAGGGAGGTGAGGC
GGCTGGGCAGGGTGGTGGGTGGCTGGAAGCGGGGAGACTCACGTATGGTGGCGGCTGGCCTTGGA
GGGAGCGAGCAGGGAGGGAGGCGAGGGGACGGACGCATCGGCGTCGGCTCATCTCCCACGAGCGC
ATAGCGTTATTGGGTGAAGAGGGGGNAG
```

sgRNA Design & Vector Construction

This selection of the sgRNA was done manually with the aid of several gRNA prediction tools, namely CC-TOP, Benchling & CRISPR-P. The predicted *OsSCE1* sgRNA was selected and inserted into virtual vectors using Benchling's Golden Gate Assembly.

RESULTS & DISCUSSION

OsSCE1 Gene prediction & Annotation

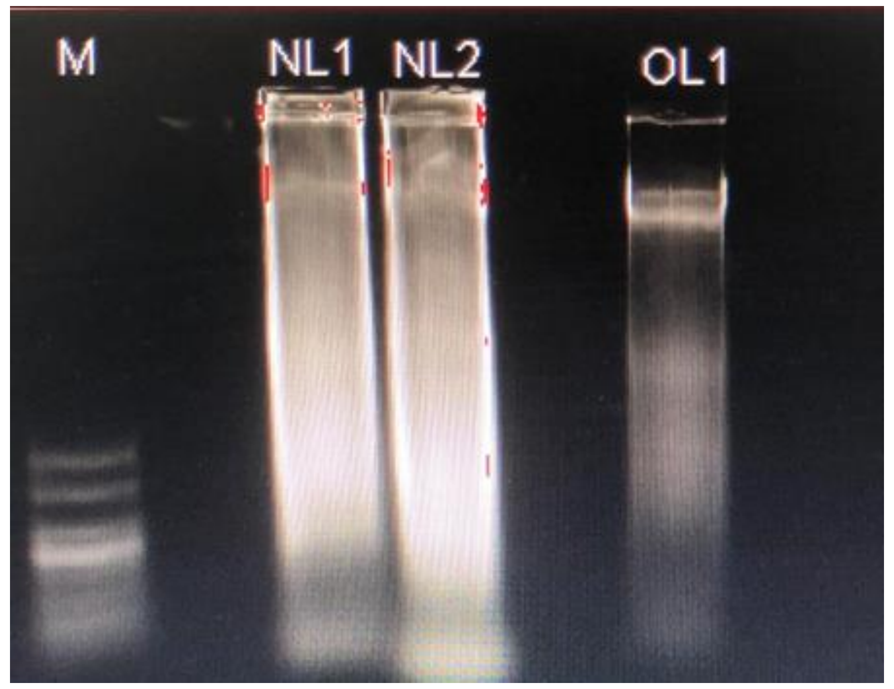
Putative OsSCE1 gene

```
>Putative OsSCE1 - Nucleotide Sequence
ATGTCCTCCCCGTCCAAGCGCGCCGAGATGGACCTCATGAAGCTCATGATGTCCGACTAC
AAGGTGGAGATGGTGAACGACGGCATGCAGGAGTTCTTCGTGGAGTTCCGCGGCCCGACC
GAGTCCATCTACCAGGGCGGCGTGTGGAAGGTGCGCGTGGAGCTCCCGGACGCCTACCCG
TACAAGTCCCCGTCCATCGGCTTCGTGAACAAGATCTACCACCCGAACGTGGACGAGATG
TCCGGCTCCGTGTGCCTCGACGTGATCAACCAGACCTGGTCCCCGATGTTCGGCGAGATC
ACCCTCGTGCTCGTGATCATCTCCACCGACCTCGTGAACGTGTTCGAGGTGTTCCTCCCG
CAGCTCCTCCTCTACCCGAACCCGTCCGACCCGCTCAACGGCGAGGCCGCCGCCCTCATG
ATGCGCGACCGCCCGGCCTACGAGCAGAAGGTGAAGGAGTACTGCGAGAAGTACGCCAAG
CCGGAGGACGCCGGCGTGACCCCGGAGGACAAGTCCTCCGACGAGGAGCTCTCCGAGGAC
GAGGACGACTCCGGCGACGACGCCATCCTCGGCAACCCGGACCCG
```

Sequence ID	Description	Length	GO IDs	GO Names	Enzyme Codes
FGENESH_25 886exon(s)	ubiquitin-conjugating enzyme E2 4	195	P:GO:0000209 P:GO:0006511 F:GO:0005524 F:GO:0061631 C:GO:0005634	P:protein polyubiquitination; P:ubiquitin dependant protein catabolic process; F: ATP binding; C: nucleus	E.C:2.3.2.23

The protein sequences from *Oryza sativa* ~~subsp. *indica*~~ for chromosome 10 were predicted and annotated and putative *OsSCE1* was obtained.

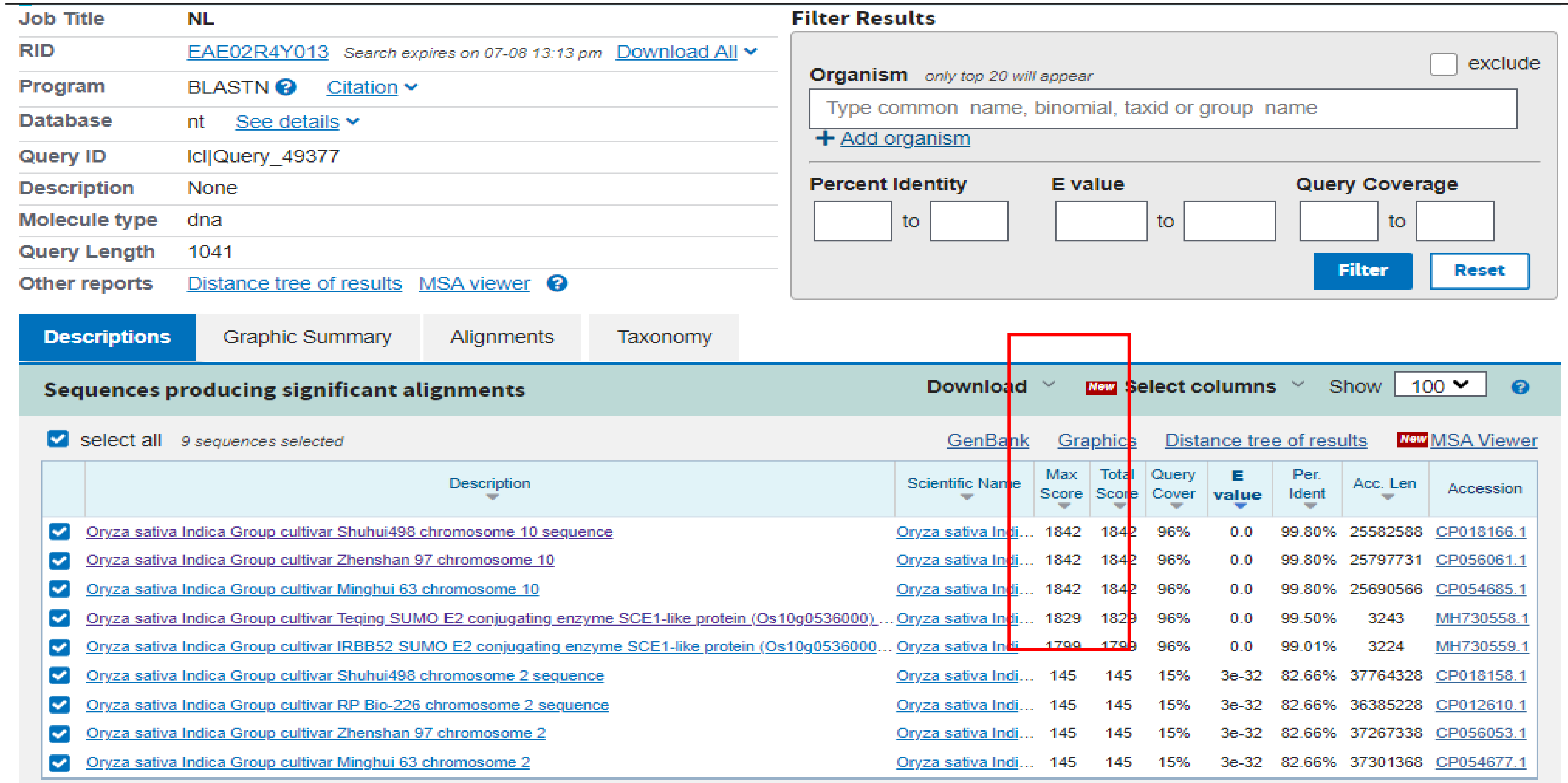
Validation of predicted *OsSCE1* gene via sequencing



pls all your result here
picture of
- your prime seq
- gel of your PCR
- your sequencing

```
>NL1.RVRS

GTAAGGAACTGAATGTTGTAAGATTTAGTTGCTGAAAGATGGATTATAGTCACCATCACTACTGT
CTTCATAATCAACATCACTTCCAAAACACTTTATCTTGCATATGGGGGATATATAGTAATCTTC
AGTGGCGGACCCAGGATTTCTTCATGCCCTAGGCCAAATCTAATTGCTACTATCCCATACCATAA
ATGAGACCCACACTAATACATAATAAGTACAATACAAATTATTATAGTTTCGAAAGGAAAAAAAC
AACAACGCAAGAAACAGCATCTAACCTCAAAATGACAGCTATTTGCTCTTTGAAAGATGCATCTC
CATCATACTCATCAACAAAAATGGCGGATTTTTTTTCCACGGAAAGCATTTAGAATGGTTTTACC
CGTCTCAATTGCACAAGATTTACACTCTAGATTTTATATTTTTAGTGAAGTGAATAGATGATTCC
CGGGAGAATTTTGAAATGTTTCTGGTCATCTTTATTCTTCTCTTTATACCCCATCAATCATCTAT
AGATTTCATCGCTAAGCCTCTTCACCTTCTATCTCTATTAAGTTTAAAACAAATTGAATGGAATC
AGAAGATCTATAGGTGATCTAGGCCATTTAGGCATTTGAGAAATGGGAATACCAGAAATAGATTA
TGAAAAACTCACCTGTCAAGGTCCTCGTCCGCTCCTTGCCCGACAGAGTGGAGGTCTGAGACTCT
GAGCACCTGCCGCTTGCCGCAGGCGCACCGTGCAGGGGGGTTGTGCAGAAAGAGCCTTGTTGCCG
GGCCGCCGAGCCAGCGAGAGGAGCTGAGGAGGCAGTGGACATGGACGGAGGGAGGGAGGTGAGGC
GGCTGGGCAGGGTGGTGGGTGGCTGGAAGCGGGGAGACTCACGTATGGTGGCGGCTGGCCTTGGA
GGGAGCGAGCAGGGAGGGAGGCGAGGGGACGGACGCATCGGCGTCGGCTCATCTCCCACGAGCGC
ATAGCGTTATTGGGTGAAGAGGGGGNAG
```



Job Title	**NL**
RID	EAE02R4Y013 *Search expires on 07-08 13:13 pm* Download All
Program	BLASTN Citation
Database	nt See details
Query ID	lcl|Query_49377
Description	None
Molecule type	dna
Query Length	1041
Other reports	Distance tree of results MSA viewer

Sequences producing significant alignments

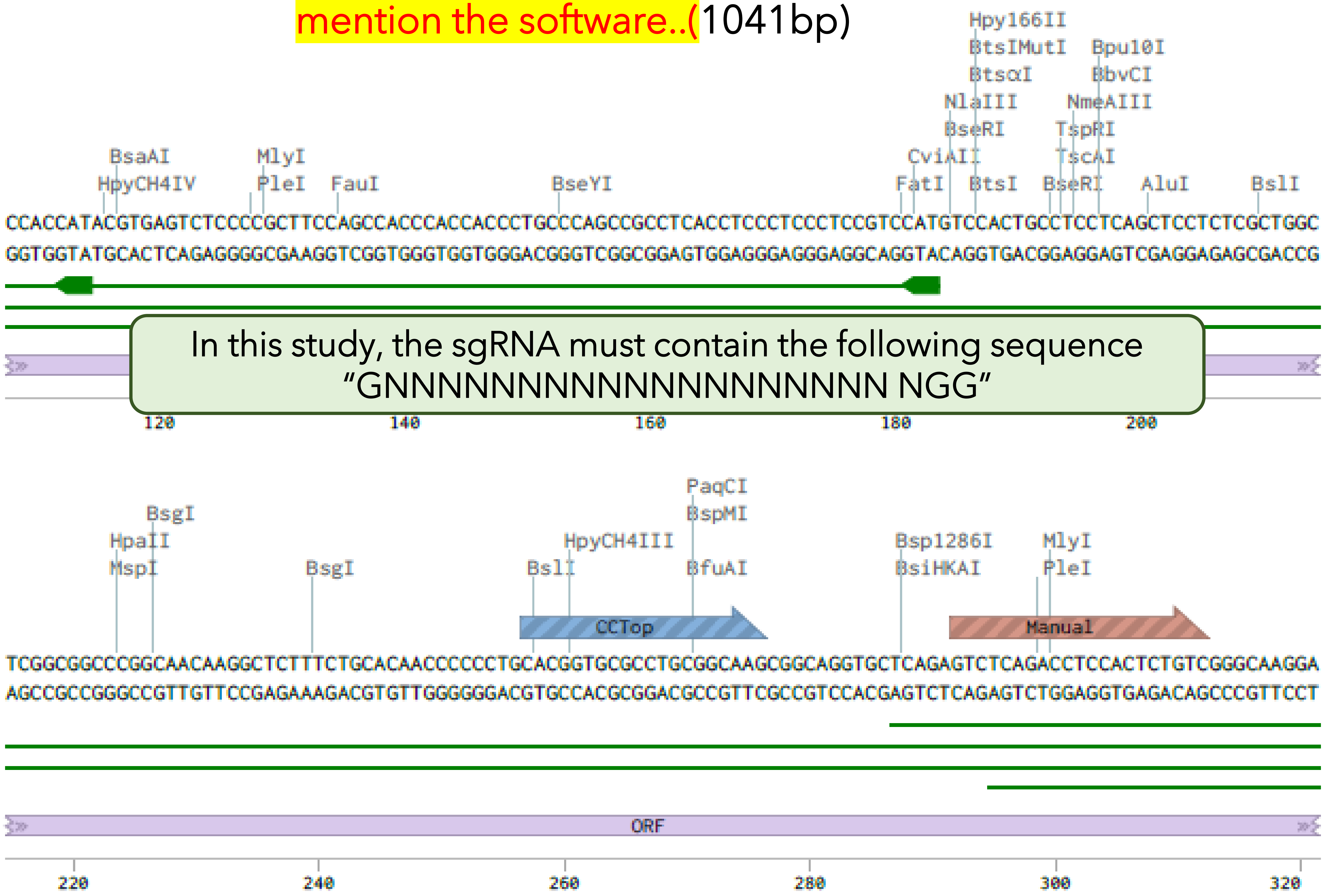
☑ select all *9 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Oryza sativa Indica Group cultivar Shuhui498 chromosome 10 sequence	Oryza sativa Indi...	1842	1842	96%	0.0	99.80%	25582588	CP018166.1
☑	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 10	Oryza sativa Indi...	1842	1842	96%	0.0	99.80%	25797731	CP056061.1
☑	Oryza sativa Indica Group cultivar Minghui 63 chromosome 10	Oryza sativa Indi...	1842	1842	96%	0.0	99.80%	25690566	CP054685.1
☑	Oryza sativa Indica Group cultivar Teqing SUMO E2 conjugating enzyme SCE1-like protein (Os10g0536000)...	Oryza sativa Indi...	1829	1829	96%	0.0	99.50%	3243	MH730558.1
☑	Oryza sativa Indica Group cultivar IRBB52 SUMO E2 conjugating enzyme SCE1-like protein (Os10g0536000...	Oryza sativa Indi...	1799	1799	96%	0.0	99.01%	3224	MH730559.1
☑	Oryza sativa Indica Group cultivar Shuhui498 chromosome 2 sequence	Oryza sativa Indi...	145	145	15%	3e-32	82.66%	37764328	CP018158.1
☑	Oryza sativa Indica Group cultivar RP Bio-226 chromosome 2 sequence	Oryza sativa Indi...	145	145	15%	3e-32	82.66%	36385228	CP012610.1
☑	Oryza sativa Indica Group cultivar Zhenshan 97 chromosome 2	Oryza sativa Indi...	145	145	15%	3e-32	82.66%	37267338	CP056053.1
☑	Oryza sativa Indica Group cultivar Minghui 63 chromosome 2	Oryza sativa Indi...	145	145	15%	3e-32	82.66%	37301368	CP054677.1

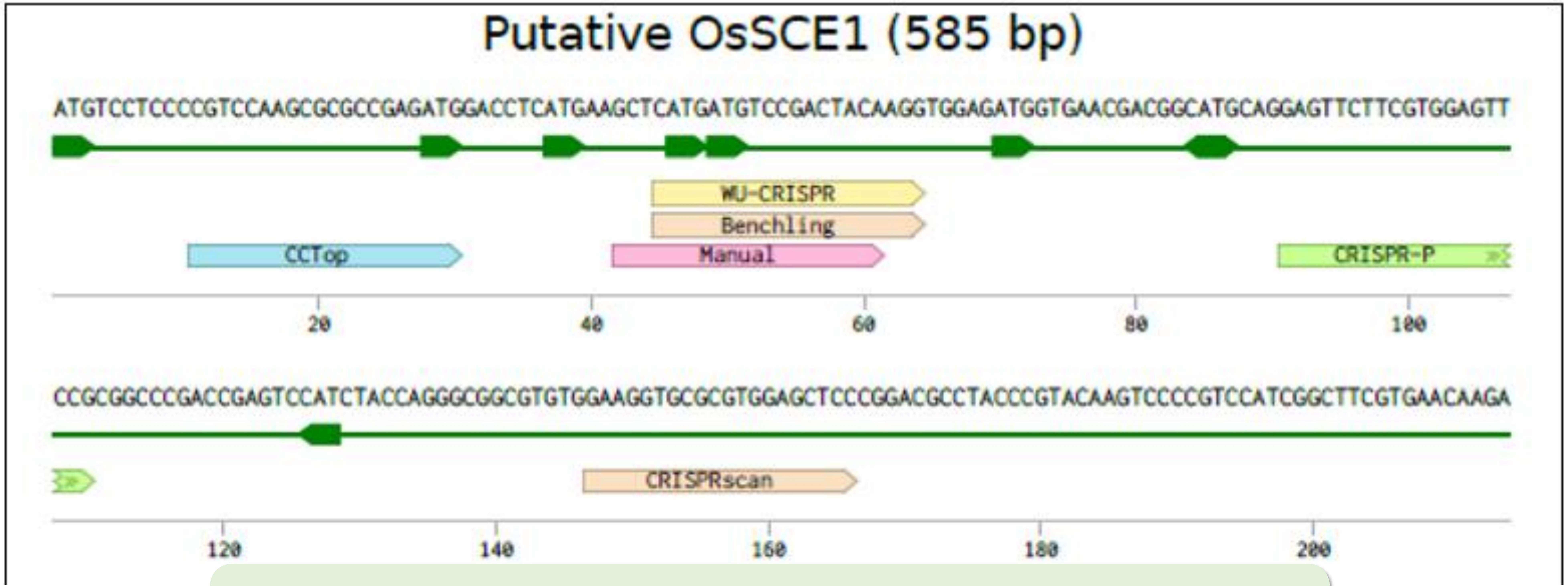
Resources: NCBI

sgRNA Design

Sequence Mapping using …. need to mention the software..(1041bp)



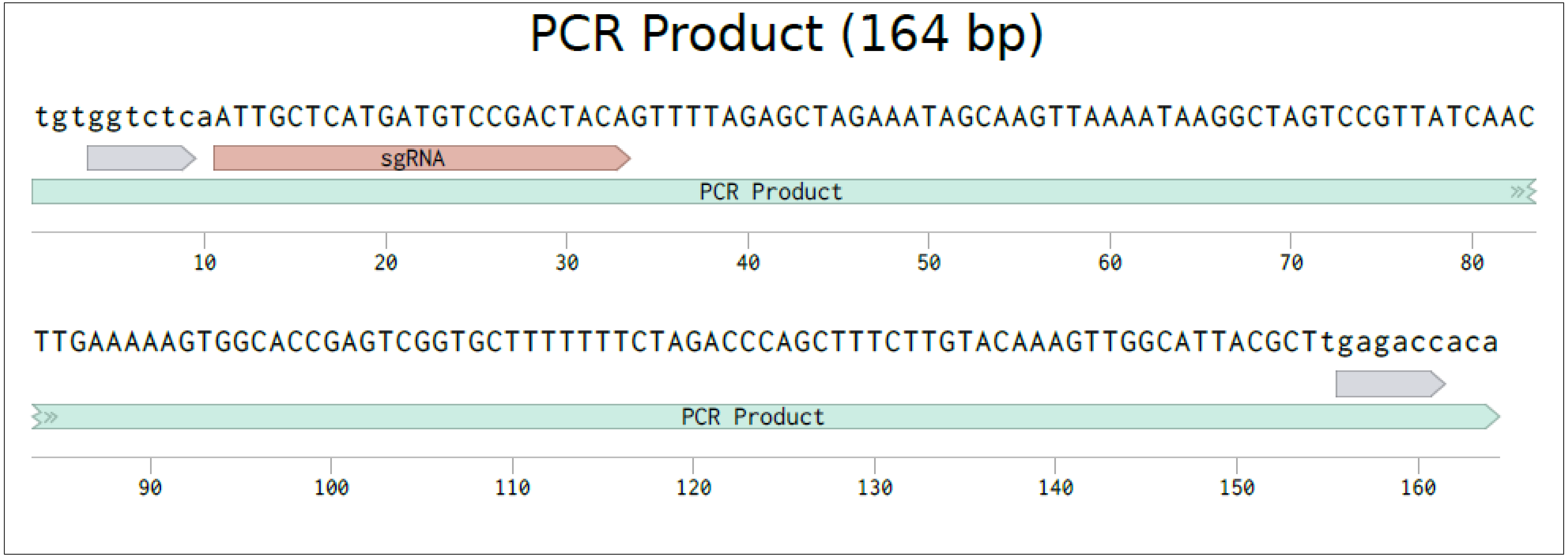
In this study, the sgRNA must contain the following sequence "GNNNNNNNNNNNNNNNNNNNN NGG"



Considering important sgRNA design criteria for plants, the *in silico* sgRNA was selected (pink colored).

CRISPR Construct: Level 0

Using PCR amplification approach, the sgRNA (Level 0) will be amplified using pICH86966::AtU6p::sgRNA_PDS construct as a template. The sgRNA was arranged in the following sequence virtually to represent the resulting PCR product.


PCR Product (164 bp)

tgtggtctcaATTGCTCATGATGTCCGACTACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC

TTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTACGCTtgagaccaca

CRISPR Construct : Level 1

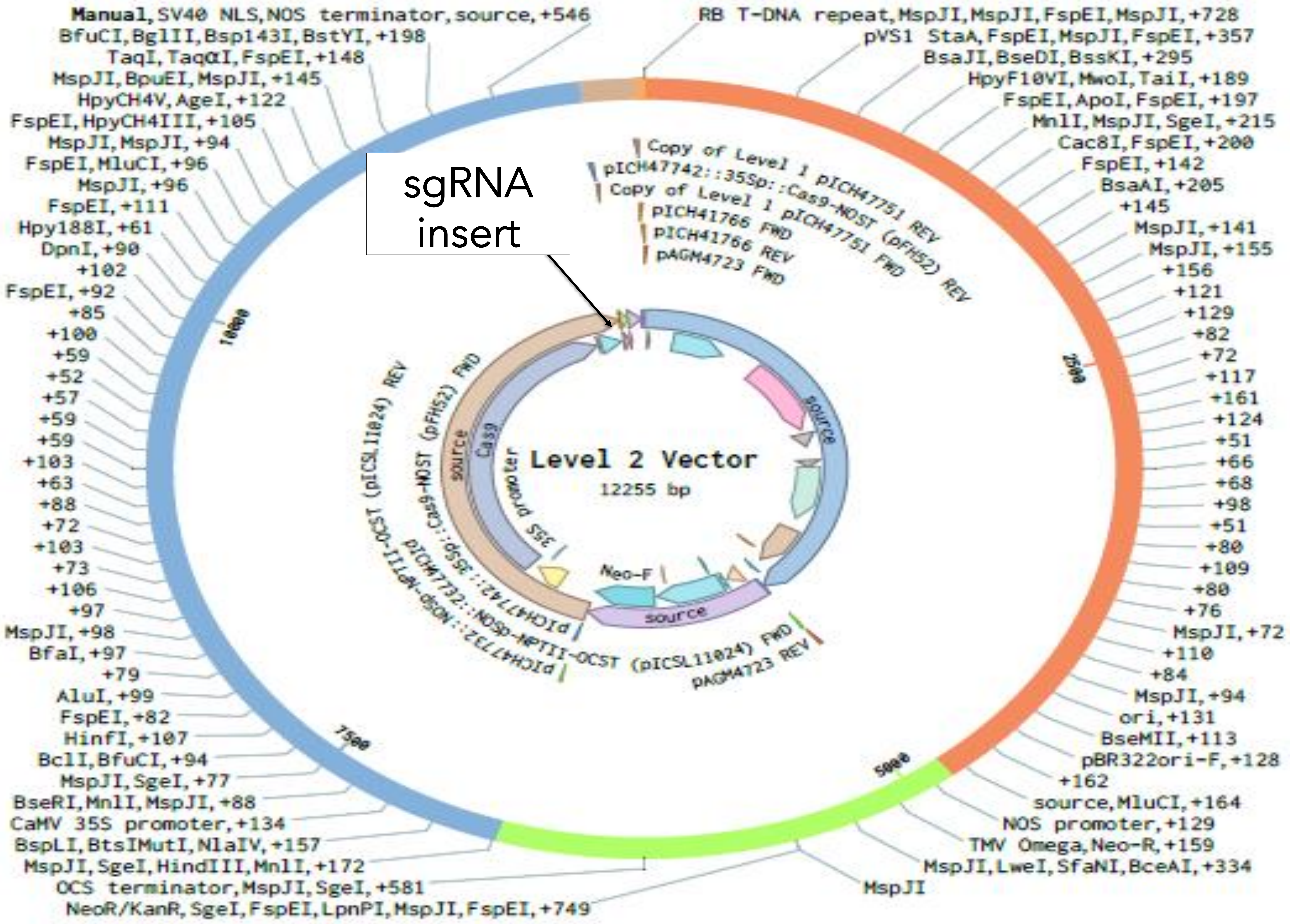
The resulting **PCR product** and vector PICSL01009 ::AtU6p (SpecR) was **delivered** to **Level 1 destination vector** pICH47 751, producing Level 1 AtU6p::sgRNA, with the sgRNA placed under the Arabidopsis U6 promoter.

Level 1 Vector (4600 bp)

CRISPR Construct: Level 2

Level 2 assembly was executed by **introducing Level 1 construct** (pICH47732::NOSp::NPTII-OCST,pICH47742::35Sp::Cas9-NOST,pICH47751::AtU6::sgRNA,pICH41766) **into Level 2 destination vector** pAGM4723. The cut-ligation reaction was done using *BpiI* (*BbsI*), producing Level 2 (NPTII-Cas9-sgRNA).

Level 2 Vector (12255 bp)



CONCLUSIONS

To conclude, the *in silico* CRISPR construct for *Oryza sativa* subsp. *indica* was developed. Using a gene associated with stress response in rice called *OsSCE1* (SUMO E2-Conjugating Enzyme), the CRISPR/Cas9 system will generate double-strand breaks in the targeted sequence. As a result, *OsSCE1* gene knockout will occur.

FUTURE WORKS

Future studies should aim to replicate and support the study through other experimental categories like *in vivo* studies.

REFERENCES

- Adnan, N., & Nordin, S. M. (2020). How COVID 19 effect Malaysian paddy industry?
- Adoption of green fertilizer a potential resolution. In *Environment, Development and Sustainability* (Issue 0123456789). Springer Netherlands. https://doi.org/10.1007/s10668- 020-00978-6
- Bernama. (2016, July 18). Nearly 6,000 rice farmers affected by drought. *The Star*. Retrieved from https://www.thestar.com.my/news/nation/2016/07/18/nearly-6000-rice-farmers- affected-by-drought
- Bi, H., & Yang, B. (2017). *Gene Editing With TALEN and CRISPR/Cas in Rice* (pp. 81–98). Bhasin, M., & Raghava, G. P. S. (2006). Computational methods in genome research. *Applied*
- *Mycology and Biotechnology*, *6*(C), 179–207. https://doi.org/10.1016/S1874- 5334(06)80011-0
- Brooks, C., Nekrasov, V., Lippman, Z. B., & Van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiology*, *166*(3), 1292–1297.
- Chamhuri, S., Mohd Idris, N., & Morshed, G. (2014). Issues and Challenges Facing Rice Production and Food Security in the Granary Areas in the East Coast Economic Region (ECER), Malaysia. *Research Journal of Applied Sciences, Engineering and Technology*, *7*, 711–722.
- Doyle, JJ; Doyle, J. (1990). Doyle&Doyle_Focus_1990_CTAB.pdf. In *Focus* (Vol. 12, Issue 1, pp. 13–15).