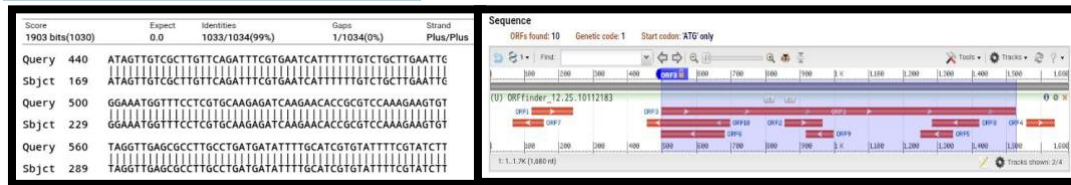# Computational Analysis of Unknown DNA Sequences Using Integrated Bioinformatics Tools

SUBMITTED BY: KASHIF NAQEEB

REG NO. 04032213041

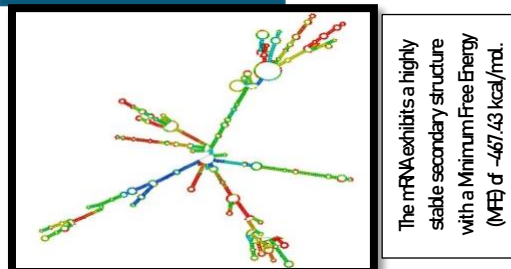## DNA Sequence Analysis

GENE: FBD-associated F-box protein

ORGANISM: Sesamum indicum (Sesame)

PROTEIN: F-box protein

Top hits confirm Sesamum indicum FBD-associated F-box protein with 99.9% identity and E-value 0.0 – GC%=38.9

ORF3 selected as the longest complete reading frame (342 aa) with valid start/stop codons

RNA Analysis

The mRNA exhibits a highly stable secondary structure with a Minimum Free Energy (MFE) of −467.43 kcal/mol.
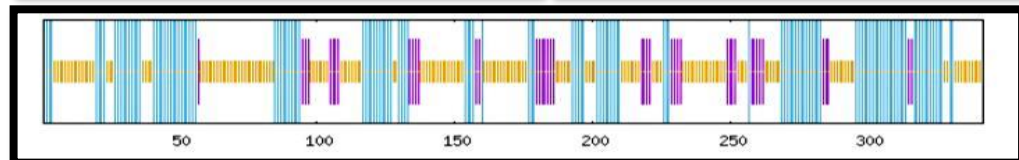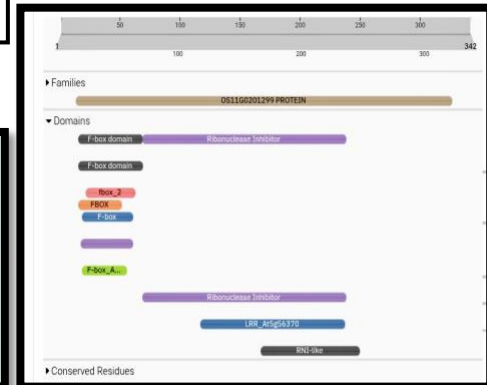
5'UTR = 1–497 Bases, 3'UTR = 1527–1680

Note: Analysis of the untranslated regions (UTRs) revealed no significant regulatory motifs, such as Kozak sequences or PolyA signals

Protein Analysis

Table of protein properties

Property	Value
Molecular Weight	39703.33 Da
Theoretical pI	8.69
Instability Index	36.35 (Stable)
Aliphatic Index	103.95

Predicted as an SCF E3 ubiquitin ligase component, its structure is dominated by random coils (48.5%) and alpha-helices (38.6%), enabling flexible protein interactions
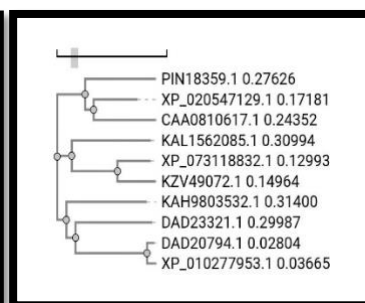
Subcellular Localization Prediction

Sequence
Predicted localizations: Nucleus
Predicted membrane association: Soluble
Predicted signals: Nuclear localization signal

Localization	Cytoplasm	Nucleus	Extracellular	Cell membrane	Mitochondrion	Plastid	Endoplasmic reticulum	Lysosome/Vacuole	Golgi apparatus	Peroxisome
Probability	0.2575	0.6426	0.0401	0.0617	0.0680	0.0484	0.1199	0.0528	0.0511	0.0080

Localization predicted as Nuclear; DeepLoc detected a specific Nuclear Localization Signal (NLS), reinforced by high confidence scores in WoLF PSORT

Phylogenetic Analysis

10 sequences

PIN18359.1
XP_020547129.1
CAA0810617.1
KAL1562085.1
XP_073118832.1
KZV49072.1
KAH9803532.1
DAD23321.1
DAD20794.1
XP_010277953.1

PIN18359.1 0.27626
XP_020547129.1 0.17181
CAA0810617.1 0.24352
KAL1562085.1 0.30994
XP_073118832.1 0.12993
KZV49072.1 0.14964
KAH9803532.1 0.31400
DAD23321.1 0.29987
DAD20794.1 0.02804
XP_010277953.1 0.03665

Phylogenetic analysis confirms the protein is evolutionarily conserved; the Sesamum indicum homolog clusters closely with PIN18359.1, indicating a recent common ancestor and shared biological function

Protein–Protein Interaction Analysis

BIOLOGICAL SIGNIFICANCE

The protein interaction network reveals that the query F-box protein (MCD7.13 homolog) is centrally integrated into the Ubiquitin-Proteasome System. Strong functional associations were identified with other F-box/kelch-repeat proteins (e.g., MGL6.21, K6M19.17), which act as substrate recognition components of SCF (Skp1-Cullin-F-box) E3 ligase complexes. This network confirms the protein's role in mediating the targeted ubiquitination and degradation of specific regulatory proteins, essential for plant development and stress signaling