



Harnessing Bioinformatics For Seed Nutritional Quality: Unraveling *Cis*-Regulatory Mechanisms In Nutrient Storage And Nitrogen Assimilation

Dr. V. K. Singh, Information Officer, Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh 221005

Abstract

The management of seed nutritional quality is crucial for improving the dietary value of staple crops. Recent studies have explored various *cis*-regulatory elements, such as the Legumin box and Prolamin-box (P-box) which play significant roles in regulating protein and nutrient storage in seeds while Nitrate-Responsive Elements (NRE) are critical for nitrogen assimilation and essential for seed protein content. This manuscript discusses key regulatory elements and their contribution to seed nutritional quality, focusing on nitrogen uptake, storage protein synthesis, and relevant genetic modifications aimed at enhancing seed quality.

Introduction

Seeds are vital for human nutrition, providing essential proteins, oils, and carbohydrates. The quality of seeds, in terms of nutrient content, is largely controlled by the expression of genes involved in storage protein synthesis and nitrogen metabolism. Regulatory mechanisms involving *cis*-acting elements are critical in this process. These elements act as binding sites for transcription factors, ensuring precise gene expression during seed development. Understanding these regulatory networks is essential for improving seed quality through genetic engineering and agricultural practices.

Keywords: Legumin Box, Prolamin-Box, Nitrate-Responsive Elements, Seed Nutritional Quality

Key *Cis*-Regulatory Elements in Seed Nutritional Quality

Legumin Box

The Legumin box is a *cis*-regulatory element that has been identified in leguminous plants, controlling the expression of legumin, a major storage protein. This element ensures that seed storage proteins are synthesized during the late stages of seed development, contributing significantly to the seed's nutritional value. The *cis*-acting analysis details for legumin box are reported from **Figure 1** to **Figure 6**.

Figure 1: FASTA Protein sequence of *Vicia faba* legumin B

>X03677.1 *Vicia faba* DNA
 CATAGGCATGATGCTGAAGAACTACACACGTTCTGTCCACACGTTACTCTCTCACTGTTCTCCTCTTCTCTATAAATCAACGCGGCCACAG
 CTTCTCCACTTCACCACCTTCACCACCTTCACTCACAATCCTTCATTAGTTGTTTACTATCACAGTCACAATGAAATGTTGTTGTTGCT
 TTCACCTTTCCTTGCTTCTCTTTACAAGCACATGTTTAGCAACTAGCTCTGAGTTTGACAGACTCAACCAATGCCGGTTAGACAACATCAATGCA
 TTGGAACCTGACCAACCGTGTGGAGTCCGAGAGCTGGGTCTCACTGAGACATGGAATCCAAATCACCTGAGCTAAGATGTGCCGGTGTGTCTCTTA
 TCAGACGCACCAATTGACCTTAATGGACTCCACTTGCCATCTTACTCAOCTCTCCACAGTTGATTATCATCATCAAGGAAAGGGTGTATTGG
 ACTTACACTTCCTGGTTGTCCCCAGACTTACCAAGAGGACGCTTCATCACAATCTAGACAAGGATCCAGGCAGCAACAACCTGACAGTCACAG
 AAGATTCCGCCGATTGAGAAAGGTGATACATTGCCATTCATCCGGAAATTCATATGTGACATATAAACAATGCCGATGAACCGCTTGTGTGCA
 TTAGTCTTCTTGACACTTCCCAACTTGCAAAACAGCTTGATTCAACCCGAAGATTAAGTATGTAACATATACATTGCTCTGTTATTA
 ATTGTTCTAGAACTTTAAATCAATCAAAATCGCATGTAAATATGTGTGTTTTCAGGTATTTTACTTGTGTTGGGAACCCGGAGGTAGATTTCCCTG
 AAACACAGGAGGAGCAGCAAGAAAGACATCAACAAAACATAGTTTACAGTTGGACGTCCGGGTGGACAGCACCAACAAGAGAGGAATCTGA
 AGAACAAGAGGACGGTAACAGCGTTTTGAGTGGCTTCAGTTTCAGAGTTTTTAGCACACACATTCAACACCCGAAGAGGATACAGCTAAGAGACTT
 CGTTCTCCACGCGCAAAAAGGAATCAAATCGTGAGAGTTGAGGGCGGTCTCCGCATTATCAACCTGAGGGGCAGCAGGAAGAAGAAGAAGAG
 AGGAGGAGGAGAAGCAGAGAAGTGACAGGGAAGGAATGGTTTGGAGAAACCATCTGTAGTCTCAAGATTCGTGAGAACATTGCTCAGCCTGC
 ACGTGCAGACCTCTATAACCCAGCTGCCGGTAGTATCAGCACCGCAACAGTTTAAACCTTCCATCTCCGCTATTACGGCTCAGTCCCGAA
 TATGTTCTGCTTTACAGGGTACGTATAGTATTAACATTTTAAACCAATTCATATATTCATATAAATGATTTTTTAATAAACTAATCAATAACAT
 GCATATGCATGTATGTATATGCAGAATGGTATATATGCTCCACACTGGAACATAAACGCAACAGCTCTGCTGTACGTGATAAGAGGAGAGGAA
 GAGTTAGGATTGTGAACTCCCAAGGAAACGAGTGTTCGACAACAAGGTGACAAAGGGACAGTTGGTGGTGGTACCACAAAACCTTTGTGGTGGC
 GGAACAAGCTGGAGAGGAAGAAGGATTAGAGTATCTTGTGTTCAAGACAAATGACAGAGCTGCTGTTAGCCACGTACAGCAGGTGTTTAGGGCC
 ACTCCTCAGATGTTCTTGCAAAATGCTTTGGTCTTCGTGAGCGCAAGTCACAGAGTTAAAGCTCAGTGGAAACCGTGGCCCTCTGGTTCAAC
 CTCAGTCT

Figure 2: FASTA CDS Nucleotide sequence of *Vicia faba* legumin B

[illegible]

Figure 3: FASTA CDS Nucleotide sequence of *Vigna unguiculata* legumin B

Figure 4: FASTA CDS Nucleotide sequence of *Phaseolus vulgaris* legumin B

X03677.1_Vicia_faba_DNA	CCATAGCCATGCGTCTGAAGAATGTCAACACGTTCTG-----TCACACGTGTT	50
ANINZ01023743.1_Phaseolus_vulgaris_DNA	CCATAGCCATGCAAGCTGAAGAATGTCTCAACTC-----ACCTTACGTGTG	46
CP039353.1_Vigna_unguiculata_DNA	CCATAGCCATGCAAGGTGAAGAATGTCCAACTCAGTAACTACCTTCACCTTACGTGTC	60

X03677.1_Vicia_faba_DNA	ACTCTCTCACTGTTCTCCTCTT-CATATAAATACCGCGCCACAGCTTCTCCACTTCACC	109
ANINZ01023743.1_Phaseolus_vulgaris_DNA	TCTCTCTCACTCTCTT-CTCTTC-CATATAAATCAACACCCCTCTC-TTCTACACTTCACC	103
CP039353.1_Vigna_unguiculata_DNA	TCTCTCTCACTCTCTT-CTCTGCAATATAAATCAACACCCCTGTG-TTCTACATTTCAACC	118

X03677.1_Vicia_faba_DNA	---ACTTCACCACCTTCACTACAATC---CTTCAATAGTTGTTTACTATCACAGTCA	162
ANINZ01023743.1_Phaseolus_vulgaris_DNA	AACCCATTCACTACTTCCTTTGCTTACTTCCTTCAAC	140
CP039353.1_Vigna_unguiculata_DNA	AACCCCTTCAACACTTCCTTTACT-----TCAAC	147

X03677.1_Vicia_faba_DNA	ATG	165
ANINZ01023743.1_Phaseolus_vulgaris_DNA	ATG	143
CP039353.1_Vigna_unguiculata_DNA	ATG	150



Figure 6: MEME motif analysis based on alignment of legumin B from *Vicia faba*, *Vigna unguiculata* and *Phaseolus vulgaris*

Prolamin-Box (P-Box)

The Prolamin-box (P-box) regulates the synthesis of prolamin proteins, which are essential for seed protein storage in cereal grains such as wheat, maize, and rice. These proteins are rich in essential amino acids and play a pivotal role in seed quality. The P-box mediates the interaction between prolamin gene promoters and specific transcription factors that activate gene expression during seed maturation. The *cis*-acting analysis details for Prolamin-box are reported from **Figure 7** to **Figure 8abc**.

>X13306.1 *Triticum aestivum* LMW glutenin-1D1 gene

```

GATCCGGGAGAGGCTGCGGCGGCGAGCGGTCCGGCCGGCGAGGAGGACGGGGAAGGGAGAGACGAAGACG
ACGCAGAATCGGCGGAATGTGGGCTGGGCTTGTGACTTTGAGGCCCTCCGAAAACTATAGCCAGTTTCGAA
TTGGTGCCCTAACACACACACACTTACGTTGGGCTAATCGCTCGCTCCCTGCCCTGCTCAAAATTTTT
TTTGCTCCAGGCTGGGGGCTCTGTTCCACCCCCATCTATCGCTCCACCTCCAAACAAAAAAATCTA
TCACTCCACCTCCACTCCAAAAATATAAAATTTCTATCAATCCACCTACGCTCGAAAAAGAAATCTATC
ACTCCACCTCAGCATTGATGTCTCTAGCTTGTAGAACTGCCATCCTTTACATGTAAAACGGATTGATG
AGTCATGTCTGCTCTATAGACGTCAGTTTCTTATCATCTTACAGGAAAGTACAAAGTTAGTTTTCTG
AAAAGCAACCGAATATAGAAGAACTCCACACTCAAGGCTTTACTAATCGAGCATATCCTAACAGCCCA
CACATGATTGCAAACTTAGTCATACACAAGTTTGGCTTTCTTGTITACGGCTGACAGCCTATACAAGGT
TCCAACTCGGTTGTAAAAGTGATACATCTTGTAAAGTGT TGACATGTAAGTTAAT AGGTGAGTCA
TATATAGCAAAATATCGGGTTTCTGTACTTTGTGTGTGATCGTATGCACAACTAAAAATCACTTTGATG
ATCAATATATCCAAAAGTACGCTTGTAGCTAGTGCAAACTTAACCAATGTAAACAAAATAATTCAATTC
GATGGAGCCAAACAGAATTATTAAGCTGATGCAAGAAGGAAAGAGGTGGTTCTCGGGCTACTATAAA
TAGGCATGAAGTATAAGATCATCAAGCAAGCATCAGAACCAGCAACACTAGTTAACACCAATCC
ACCATGAAGACCTTCTCTGCTTT GCCCTCTCGCGTGTGGGCGACAAGTGCAATTGCGCAGATGGAGA
CTAGATGCATCCCTGGTTTGGAGAGACCATGGCAGCAGCAACCAATACCACCAACAGACATTTCCACA
ACAACCACTATTTTCAACAACAACAACAACAACCACTATTTCTCAACAACCATCATTTTCGAGCAACAA
CCACCATTTTGGCAGCAACAACCAACCAATTTCTCAGCAACAACCAATTTCAACCAAGCAACCAACATTTT
CGCAGCAACAACAACCTAGTTTCTACCGCAACAACCAACCAATTTTACAGCAACAACAACCAAGTTTACCTCC
ACAACATCACTTTTCCACAACAACAACAACAACAACCAACAGCTGGTGCAACAACAATCCCTGTTGTT
CAGCCATCCATTTGCGAGCAGCTAAACCCATGCAAGGTATTCCTCCAGCAGCAGTGCAGCCCTGTGGCAA
TGCCCAACGTCCTTGTAGGTGCGAAATGTTGCGAGCAGCAGTGGCCATGTGATGCAACAACAATGTTG
CCAGCAGTTGCGGCAATCCCCAGCAATCCCGCTATGAGGCAATCCGTGCTATCTACTCCATCATC
CTGCAAGAACAACAACAGGTTCCATCCAATCTCAGCAGCAGCAACCCCAACAGTTGGGCAAT
GTGTTTCCCAACCCCAACAGCAGTGCAGCAGCAACTCGGGCAACAACCTCAACAACAACATTTGGCACA
GGGTACCTTTTGCAGCCACACAGATAGCTCAGCTGAGGTGATGACTTCCATTGCGCTCCGATCCTG
CCAACGATGTGCAGTGTAAATGTGCGGTTGTACAGAACCACCACTAGTGTGCCATTGCGGCTTGGCACCG
GAG TTGGTGCTCTACTGATAAGGAAAGATCTCTAGTAATATATAATTGGGTACCGTTGTTTAGTCGATGG
ATATGTCGATGCAGCGGTGACAAATAAAGTGTACACAATGTCTGTGTGACCGGCCAACTAGTTGTT
TAAATTCTGAATAAAATAAAATAAAGTTGTATCAAGACAATGTTCAATTGGCATTGTGTGGATGTCA
ATCTGATTGCCATGTCTGCAAGTTTGTCTTTTCTTGTGCAAGCGCAACCTGGTGCCTTAAT
TAATTATCAATGTACTGGAATAATCACTATTTAAATATAATAGTGTCACTGTAAAAATTTGGGTTGAACCTC
TTTATTGGTTGGAGATTTGAGATCTTGTTTTTATTGGTTTGTATCTAGTACCACTAGTACCACTGTCT
ACAGTGACAATCACAACCAAGTTTGTGCTAATTATGTTTGTCTTCTGCAAAATACATTCAATTGATTCTT
TGCATGTCAAATTAATTTGCCGAGAAAAATACCAGCTTTTCTTGAACCTCACAACACCACTAGCTAGTT
AATTTGAAATAAGCTGCCAATTTATTTCTTGTATTTATGCCCCTTTACATCGAACCACTACAATCATTGT
GCTAGCTAGAGAAAAGGAGTTGCTGGCAGTTTCTGTCTAGAACTAAATCAGACTGAATTGAACCTGTTT
TTTACAACCTAGCTAGGAGCAGAAAACTTACTGTGAATTGCACTTGATTATATTTGATTTTGTAAAGAGC
TACAGGGAAAAATCATCAAAAAATGCCATAAAATATCTTATCGTAAAAATAGTTTACCAGGAGATACAGT
CGCAAGAAAAAAAGGCTCACCAGGAGATAGATGCTGCACATTATCATTTAGAAATTAGACCAAGAGAAA
TGTGGATTGGCGGCTCATAATCTATAAGACTATGAACAGGTTGTCCAATCATTTAGAAATTAGATCAAGA
AAATGTGGAACATAACATATTAATGGCTAGGATCATATTTGCACTGTGGGAACCTTGGTATGCGCGCTCCA
ACATGCCGTTGAAAGCATTTTCTCTGTATGCACATATGAAATATGACTGGTGGAAACAGAAGAAAC
AGTGAGATAATATGGATGCTTGGATGTTGTAACAAACATAAACTTATTATATATATATCCAAATCCACCGA
GCGGACGAACATCCTAGGAGCTTTCTCTCGGAGATCGAAAAAGAACCAATGTCCAGCCAACTGCTTC
TTCTCCATCTGATAA

```

Prolamin-box (P-box)

Figure 7: FASTA CDS Nucleotide sequence of *Triticum aestivum* LMW glutenin-1D1

>JADDXV010000724.1:113478-115955 Sorghum bicolor cultivar Rio
 chromosome_2_contig_423, whole genome shotgun sequence
 TGCTTTGGTTTTTATCTTAACCTTTCTTTGTTCCCTTGAAOCCCCACTATTGAGGATTATATTTGTAAGTTG
 TAAATAGTTTTTATCAATAAAATTTCTAGTAGGGGGTAAAACCGTCATGTTTCCTCAAAAAAATTGCCACCA
 AATTGGGAGTAATCTTCAACTGATAAAAGATTGGGCACCCATGTTTCCATCACCTAGTACAACTCCAAAG
 TGAGATATAAAAAGAAGCTAGTGATAACTTAGCTCATATTAACATAAACTAAATTTTAGCACTTCAAAATA
 GCATATCCTTTTCAAGGTGAGAGTCAAGTTCAGATCACACTGTTAGTGTGTAACAAAACACTTATCCAC
 ATGTGGTGGATCACGTCCACTTGATAATGCTATCATCCCTAGTTGGGCTACCATAACTATTTATGACACA
 TAAAGCAAAAGCATCCAAGCTCTATCCACAAATGGATGTTATCTTAACAATTACATGATATTACATACT
 AAATTACTTGAGAGCATTGGTATGCAATCCTTTATGTTTACACAAATTAGTGAGCTGCCTCAAGTTGTA
 TATCGATCCATAGTATTCACAACTCATGCAACACTACAATAAGTAACACAATATTATAGAGTTATTT
 TTATGAAAAGCAAGAAAAGGACAATAATCCCTGACATGTAAGTGAATCTTATTAGATTACCTAACCCAA
 CACAAACAAACAAATTAGTTGCACAAAGGTACAAAATCCACATCAACCTTAATCAATACGTAAAGTG
 AGTAATGAGTCACATTATCCAATAAAATATCTATCAAAGTAGTATCGTGCAAGCATTGAGATCCCATAAA
 TGCAAGAAACTCTACCATCAAAGCTATAAGTAATTCCTCATGTGTTTGCTTCATTTCATCACAACATCC
 TTCTATTGCTACCATCAATAACCTTCTGCTTAGTAGTATAGGTACCAATCAGATCTGCAACATCAAT
 ATTATTGAAATCCAACAACCCTGAGAGGCTGCTTGGTTTTGCTTTTCATGGCTCTAGTATCGAGGCG
 TGCTTGCAACACAGATAGGCGGTTGCACCTGTGGTCCACAACAAAGCCATGAACAACAACATCATCCAAA
 AACAGCAACATCAACCACCAACCAACCAACAAACAACAACAACAGCATCACCACCAACCCACAGC
 ACCAACCAAAAAAGTTCACATGCAGCAGAAACAACATCACCACAAACAAGATGTTCTGTTTATCAGCA
 ACRAAGAGTCCATGTGAACCAACAACCAAGCAGCACCACCAACAACATCAATGTGAAGGCCAGCAACAGCAT
 CACCAACAATCACAAGGCCATGGGCAACAACAAGGACAGGCCATGAGAAACAACATCAACAACAATTCC
 AGGGTCATGAGAAGCCACAACAACCAACCAAGCAACAATGTGAGGTTTATGTGCACCAACAGCCACCGCA
 ACATCAACAATGTCAACAGGGCCAGGAGCAACTACAACAGCATCAATGTCTATTGCCAAGAGCAGCCATTG
 CAACAACCAAGCACTGCCAAAAAGGGTGTGAGCAGCAACAACCTACAAGGTGCAGTTCTAGCTACTGTA
 GTGGTAGCATAAATCTAAAAAATTTGCCATGAATTCCTCAGGCAACAGTGCAACCCTTTGGTAATGCCTTT
 TCTCCATCATTGTTGTTGTAACAACCAAGCAACTGCCAAGTGTACAAACAGCAATGTTGTACAGAGCTTAGG
 CAGATGAGCCACAATACCTTCAACAGCGATCTATAGCATGACTCAATCCATAATCCAACAACAACAAC
 AACCAOCCATGTGAGTTTTGTGGAGCTCACTAGCTACTCAAAGTGAGGTGGCGATATTGACGGCAGCGCA
 GTACCTACCATCAATGTGTGGCTTGTACCACTCATGCTACCAAAAACAATCCATGTAGCAGTAATGATGCT
 TGGGTTGTCACAGTTTATTATTGTGGCTAGCTAGCCGTTATAGTCTTATAACGGTGTAAAGCAATAAA
 ATGCCATACATCATGGTGTGTTGGACAACGATATCTAAATAAGAATACTATCTATAATGTGGCATGTGCT
 GTGGTGTCAACAACCTAGTCTTTGTGGTCATAGTAGATAGGGTAGTTTGTGTTGAGCATGGTTGCCAT
 ATTAGTCAGTGTGTGGACATATGTGGCATGTGCTATGGTGTGAGCGACCTGGTCTTAGTGGTCACGATA
 GATAGGGTACTTTGTTGTTGAGGCATTGGCTCCTAATTGTAATCCTATTTCTATGTACATAATCTCAGCA
 TCTAAGCATGTAGTTGGCCACAATCTCATGATCAGCCTCTAGTGCTTTGGCACCTCGATATCTACCGTGG
 CATGTGCATGTGGTGTAGCAACCTAGTCTTAGTGGTCACAACAGATAGGGTAGTTTCTGTTGGGCCAT
 TGGCTCCTAATGTAATCATATTTCTAT

→ Prolamin-box
(P-box)

Figure 8a: FASTA CDS Nucleotide sequence of *Sorghum bicolor* LMW glutenin-1D1



Figure 8b: Protein Family analysis of *Sorghum bicolor* LMW glutenin-1D1

Co-occurring Term	PR	S%	#Together	#Compared
GO:0045735    nutrient reservoir activity	10,050.97	100.00	16,306	16,306
GO:0010431    seed maturation	231.35	0.11	18	782
GO:0043245    extraorganismal space	8,375.81	0.06	10	12
GO:0071215    cellular response to abscisic acid stimulus	243.45	0.04	7	289
GO:0048316    seed development	60.60	0.04	7	1,161
GO:0070207    protein homotrimerization	63.94	0.03	5	786
GO:0043181    vacuolar sequestering	758.56	0.02	4	53
GO:0032119    sequestering of zinc ion	321.63	0.02	4	125
GO:0042735    protein body	3,769.12	0.02	3	8
GO:0034214    protein hexamerization	21.60	0.02	3	1,396

Figure 8c: Gene Ontology details of *Sorghum bicolor* LMW glutenin-1D1

Nitrate-Responsive Elements (NRE)

Nitrate-responsive elements (NRE) are critical for nitrogen assimilation, an essential process for seed protein content. These elements regulate genes involved in nitrogen metabolism, which affects the accumulation of nitrogen-rich compounds such as proteins in seeds. Recent studies have identified specific NRE sequences in various crops, such as *Solanum lycopersicum*, where these elements contribute to nitrogen uptake and protein synthesis. The *cis*-acting analysis details for Nitrate-responsive elements are reported from **Figure 9** to **Figure 12abc**.

TTTAGGAAAGATGTTACCTTCATAAAATATAGTATCTTTTTCACGCTTCCTCATAGTTTGGCTTACAAACCCTAATCTCTTTTGT
AATGCAGCAGCCGTGTTTATTGATCTTGTGTTAATCTCTCTTCTGCTCTTCTTATTGTTCTTCTTCTTCACTATAGCGAAAAGCG
TGATTCTTCTTTGTTCATCATCATCAGCTAATCATATTGCTGTGACCCCTCAATTGGCTTCTCCACAACCCCTACACTCAAGAAGCG
ACCATCATGAAACCCCTTTTCTTCTCAAGAAATGGCATCCCAACACCTTCTCAGTCTCATGACCCATCAGATGTGCACATCAT
AAAAAAGCTTCTTCAGAAGGCACTCATCTTTATGGATCAACCCAGCATCAACGACTACAAGCAGGCCCTCTTACATGGAAAACT
TAAGAATTCACTAGAAAGACTCGTCAAAGAAGCTGAAGCAAAAACTGTGTGGTTCTGGAATTGTACCATGGGCTGCTTTCGCG
AGATGCCACAAGGAGCTGAATATTGTCGAAAGCTTCAAAATGATCACTTTGCGAAGCAATCTGGAGCTCGGTATGAAATCC
ACCTTCTAAAGAACCATAACCTTCTAGGATTCAAAAGAATACCTTTGATGAAGCATGTTGTGGCAGTGGTATGGTCTGTCTCAAG
TTCTGATAGGGCTTCTTATATACATTGGGATGGATTCACTGACCCCTGAGGCTTTGGAGAACCCTGATTGATACACTCTTCACG
CGACGATTAGAAGGCATCATTCAGGAACAATGGAGGAGTCAAGACATGCTTCTATTGTGAGGCAATTACTCTTCTGCTAAG
AAAGAGTAGGCCCTGAGTGATTATATATATGTACAGGGGAATGCCCTCTTAGCTCTGTGAACAACTAACTATACTTTGAGTT
GATGATTTAAATGCTTATTAAGTACCCTGTTTATCTTGTGTGTATCTTCAACCTTATTCCTTAAGCAGAGATTAACAACCTAAG
ATGTTGGATAACTCAAATGCCCTCTTCACTCAACTCAACTGTGATTAACCAAACCTGGGTGCTCAAATTAATCTATTGAATCTCT
CTTACCATCAAGGTTCCATAATATCTATCTAAGATGACGTAAAGAGCTGAACACAATGTCACTG

Identified *Solanum lycopersicum* NIR (Nitrite Reductase) gene

```
>FGENESH:[mRNA] 1 4 exon (s) 1113 - 3568 1764 bp, chain +
ATGGCATCTTTTTCTATCAAATTTTGGCACCTTCATTGCCAAATCCAACTAGATTTTCC
AAGAGTAGTATTGTCAAGCTCAATGCAACTCCGCCGAGACAGTGGCTGCGGCGGGGCTT
CCAGAGGTTGCTGCTGAGAGACTAGAACCAAGAGTTGAGGAAAAAGATGGATATTGGATA
CTAAAAGAGCAGTTTAGGCAAGGAATTAATCCTCAAGAGAAGGTGAAGATTGAGAAGGAA
CCTATGAAGTTGTTTCATGGAAAAATGGTATTGAGGAGTTAGCTAAGATTCCAATTGAAGAG
ATAGATCAATCAAAGCTTACTAAGGATGACATTGATGTTAGGCTCAAGTGGCTTGGCTC
TTCCATAGGAGAAAGAATCAATATGGGAGATTCATGATGAGGTTGAACTTCCAATGGA
GTAACAACAAGTGCTCAGACTCGATATTTGGCGAGTGTGATAAGGAAATATGGAGAGGAA
GGATGTGCTGATATTACGACAAGGCAAAATTGGCAGATTCGTGGAGTAGTGCTGCCTGAT
GTGCTTGAGATTCTAAAGGGACTTGAAGAAGTTGGCTTGACTAGTTTGAGAGTGGCATG
GATAATGTGAGGAATCCAGTTGGAAATCCTCTGGCTGGAATTGATCCTGAAGAAATAGTT
GACACAAGAGCTTACATAATTTGCTCTCCCAATTTATCACTGGTAATTCACGAGGCAAT
CCGGCTGTTTCTAAGTTGCAAGGAAGTGAATCCGTGTGTAGTAGGGTCTCATGATCTT
TATGAGCACCTCATATCAATGATCTTGATACATGCCTGCCATAAAAGATGGACGATTT
GGATTCAACCTGCTTGTGGGAGGGTTCTTCAGTGCCAAAAGATGTGATGAGGCAATTCCT
CTTGATGTCATGGGTTCAGCCGATGATGTTGTTCCGGTTTGCAAGGCAATACTGGAAGCT
TTTAGAGACCTTGGGTTTCAAGGGAACAGGCAGAAAGTGTAGAATGATGTGGTTGATCGAT
GAACTGGGTGTAGAAGGATTACAGGCGAGAGTTCGTAAAGAGAATGCCTCAGCAAGAGCTA
GAGAGAGCATCTCCGGAAGACTTGGTTTCAAGAAATGGGAAAGAAGAGATTATCTTGGT
GTACATCCACAGAAACAGGAAGGCTATAGCTTTATTGGTCTTACATTCCAGTGGGTCTG
GTCCAGCAGACGAGATGGATGATCTAGCTCGTTTGGCTGATGAGTACGGCTCAGGAGAG
CTACGGCTGACTGTGGAACAGAACATTATTATTTCCCAACATTGAGAAGTCAAAGATTGAC
GCACTGCTAAAAGAGCCTATTTTGAAGCAATTTTACCTGATCCACCTATTCTCATGAAA
GGTTTAGTGGCTTGTACTGGTAACAGTTTTTGTGGACAAGCCATTATTGAAACGAAAGCT
CGTTCCCTGAAGATCACCAGAGGTTCAAAGGCAAGTATCTCTAACGAGGCCAGTAAGG
ATGCACTGGACAGGCTGCCCAATACGTGTGCACAAGTTCAAGTTGCAGACATTGGATTC
ATGGGATGCTGACTAGAGATAAAGACAAGAAGACTGTGGAAGGCGCGATGTTTTCTTA
GGAGGCAGAAATAGGAGTGACTCACATTTGGGTGAAGTATACAAGAAGGCAGTTCCTTGT
GATGAATTAGTACCATTATTGTGGACTTACTTATTAAGAAGTTTGGTGCAGTTCCACGA
GAAAGAGAAGAAACAGAAGATTAA
```

Figure 10: Identified FASTA CDS of NIR gene from *Solanum lycopersicum*

```
>FGENESH: 1 4 exon (s) 1113 - 3568 587 aa, chain +
MASFSIKFLAPSLPNPTRFSKSSIVKLNATPPQTVAAGPPEVAAERLEPRVEEKDGYWILKEQFRQGIN
PQEKVKIEKEPKMLFMENGIEELAKIPIEEIDQSKLTKDDIDVRLKWLGLFHRKNQYGRFMMRLKLPNG
VTTSAQTRYLASVIRKYGEEGCADITTRQNWQIRGVVLPDVP EILKGLEEVGLTSLQSGMDNVRNPVGNP
LAGIDPEEIVDTRPYTNLLSQFITGNSRGNPAVSNLPRKWNPCVVGSHDLYEHPHINDLAYMPA IKDGRF
GFNLLVGGFFSAKRCDEAIPLDAWVPADDVVPVCKAILEAFRDLGFRGNRQKCRMMWLI DELGVEGFRAE
VVKRMPQQELERASPEDLVQKQWERRDYLG VHPQKQEGYSFIGLHIPVGRVQADDMDLRLADEYGSGE
LRLTVEQNI IIPNIENSKIDALLKEPILSKFSPDPPI LMKGLVACTGNQFCGQAI IETKARSLKITEEVQ
RQVSLTRPVRMHWTGCPNTCAQVQVADIGFMGCLTRDKDKKTVEGADVFLGGRI GSDSHLGEVYKKA VPC
DELVPLIVDLLIKNF GAVPREREETED
```

Figure 11: Identified FASTA Protein of NIR gene from *Solanum lycopersicum*

USERSEQ1 (587 aa)

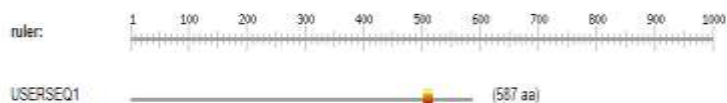
NASFSINFLAPSLMPTRFSKSSIVKLVATPPQTVAAAPPEVIAERLEPRVEKDGWILKEQFR
 QGIDPQEKVKIEKPMKLNWEGIEELAKTPIEEIDQSKLTQDDIVLKLGLFHRKMQKGRPH
 NRKLPLNGVTTSAQTRYLASVIRKYGEEGCADITTRQMQIRGVLPDPPEELKGLLEEVLTSLQS
 GNDINRNPVNPILAGIDPEEIVOTRPTNLLSQFTTGNRGRFVSVNLPRKMPGVGSHLYEHP
 HDIDLAVYPAJNDGRFRILLVGGFFSAKRCDEAIPLDHVPADQVVPVCKATLEAFRLGFRGR
 QKCRNMLIDELGVGGFAEVVKNPQQLERAGPELVQVQVRRDYLDHHPQVQEGYSFGLHI
 PVGRVQADMDLRLADEYGGELRLTVEQNIIPNIEKSDALKEPTLSKPSPPPIPLKGL
 VACTGQFCQAIETKARSLKITEEVQVQVSLTRPVNHITSCPHTCAGQVQADTGRVQCLTRON
 DQKTYEGADFLSGRIGSDSHUGVYKAVPCDELVLIVOLLINFGAVPREETED

Legend:

disulfide bridge active site other 'ranges' other sites

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
 For more information about how these graphical representations are constructed, go to <https://prosite.expasy.org/mydomains/>.

hits by patterns: [1 hit (by 1 pattern) on 1 sequence]



PS00365 NIR_SIR Nitrite and sulfite reductases iron-sulfur/scheme-binding site

504 - 520: [confidence level: (0)] TGCprrtCaqvvaD1GF

Figure 12: Functional Signature analysis of NIR Protein

>Arabidopsis

AGAAACAACTTGACCUTTACATTGCTCAAGAGCTCATCTCTT

>Spinach

ACATACAAAATGACCCTTAACCATGTCCAAGAGTCCCCTCTTA

>Bean

GAGGAGGAAGTGACCCTTGGTAGCACCGAAGAGTCCGAGGAAG

>Birch

TGTTTTTTAGGGACCTTTGGGGGTGCCAAGAGTCAAGTTATA

>Tobacco

GTAATATTAACGACCUTTGAATAACCAAGAGTCCACTAACG

>Rice1

AGCCAGCATCTGCCCTTTTGAATTCGCCAAGGGCCGCATCTGC

>Rice2

TCTCAGCATCTGCCCTTTTGAATTCGCCAAGAGCCAGCATCTG

>Maize

TGCGTCTGCCTGCCCCCTTGAAGTCCCCAAGAGCCAAGAACCA

>Sorghum

TGCATCTGCCTGCCCTTTTGAAGTCCCCAAGAGCCAATAACCA

>Brachypodium

GCGCGGCATCTGCCCTCTTGAAGTCCCCAAGAGCCAAAACTC

(a)

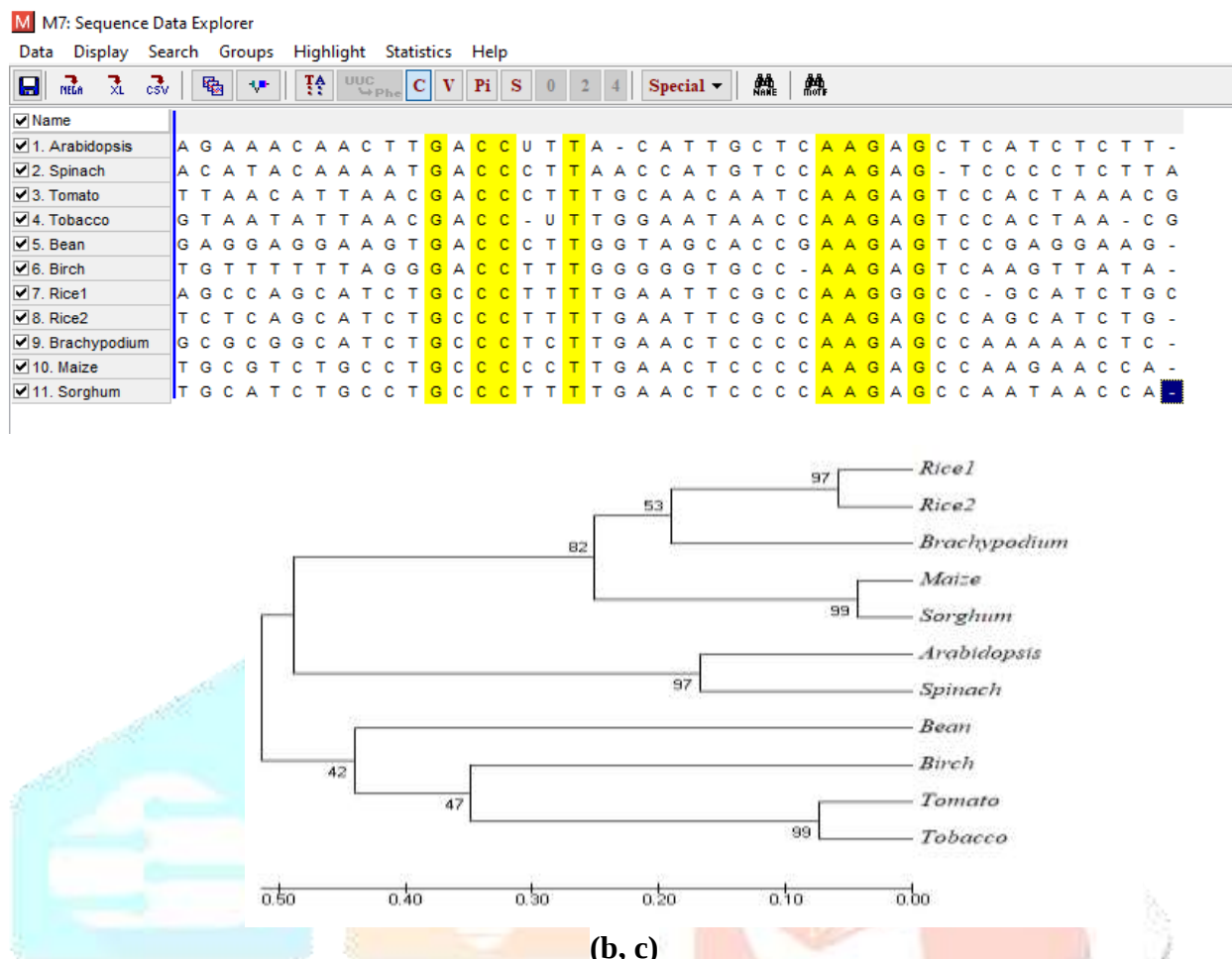


Figure 12: (a) NRE Sequences used for alignment. (b) Multiple Sequence Alignment (c) NRE element comparison Tree from 11 crops including identified Tomato NRE

Roles of Nitrogen in Seed Nutritional Quality

Nitrogen is a critical nutrient that impacts the protein content of seeds. The transcriptional regulation of nitrogen-responsive genes is mediated by NREs, which modulate the expression of nitrate reductase and nitrite reductase, key enzymes in nitrogen assimilation pathways. According to Konishi and Yanagisawa (2011), these elements are essential for coordinating the nitrogen uptake and utilization necessary for the biosynthesis of amino acids and storage proteins in seeds.

Nitrogen Assimilation and Seed Protein Synthesis

Nitrogen assimilation pathways, regulated by NREs, significantly influence the protein content in seeds. Genes such as NIR1 (Nitrite Reductase 1) are activated by nitrogen-responsive transcription factors, enhancing nitrogen utilization for protein synthesis. These regulatory mechanisms contribute to higher seed quality in crops like *Arabidopsis thaliana* and *Spinacia oleracea*.

Genetic Engineering for Enhanced Seed Nutritional Quality

Recent advances in biotechnology have enabled the manipulation of *cis*-regulatory elements to improve seed quality. For instance, enhancing the activity of the Legumin box and P-box has been shown to increase storage protein content in legumes and cereals. Additionally, the introduction of synthetic NRE sequences has improved nitrogen utilization efficiency in crops, resulting in higher protein yields.

Case Study: NIR Gene in *Solanum lycopersicum*

In *Solanum lycopersicum* (tomato), the NIR gene, regulated by NREs, plays a crucial role in nitrogen metabolism and seed protein content. By modifying the promoter region of the NIR gene to include more responsive *cis*-elements, researchers have increased the seed's nitrogen use efficiency and protein accumulation. NRE barcode is represented in **Figure 13**.

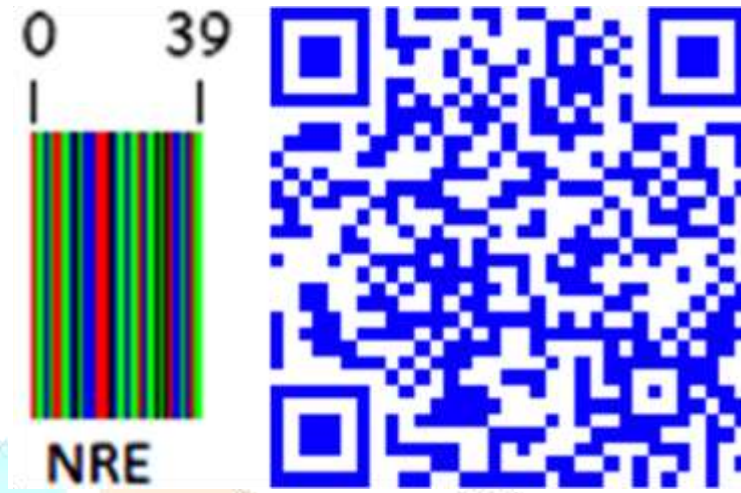


Figure 13: *Solanum lycopersicum* NRE

Conclusion

The management of seed nutritional quality can be significantly improved through the understanding and manipulation of *cis*-regulatory elements such as the Legumin box, P-box, and NREs. These elements are central to the regulation of protein and nutrient storage in seeds. Future research and genetic engineering efforts aimed at enhancing these regulatory mechanisms hold great promise for improving the nutritional value of staple crops.

References:

- Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D., & Wobus, U. (1992). *Cis*-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *The Plant journal: for cell and molecular biology*, 2(2), 233–239.
- Bäumlein, H., Wobus, U., Pustell, J., & Kafatos, F. C. (1986). The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucleic acids research*, 14(6), 2707–2720. <https://doi.org/10.1093/nar/14.6.2707>
- Colot, V., Bartels, D., Thompson, R., & Flavell, R. (1989). Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. *Molecular & general genetics: MGG*, 216(1), 81–90. <https://doi.org/10.1007/BF00332234>
- Hammond-Kosack, M. C., Holdsworth, M. J., & Bevan, M. W. (1993). In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *The EMBO journal*, 12(2), 545–554. <https://doi.org/10.1002/j.1460-2075.1993.tb05686.x>
- Konishi, M., & Yanagisawa, S. (2010). Identification of a nitrate-responsive *cis*-element in the *Arabidopsis* NIR1 promoter defines the presence of multiple *cis*-regulatory elements for nitrogen response. *The Plant journal: for cell and molecular biology*, 63(2), 269–282. <https://doi.org/10.1111/j.1365-3113X.2010.04239.x>

Konishi, M., & Yanagisawa, S. (2011). Roles of the transcriptional regulation mediated by the nitrate-responsive cis-element in higher plants. *Biochemical and biophysical research communications*, 411(4), 708–713. <https://doi.org/10.1016/j.bbrc.2011.07.008>

Vicente-Carbajosa, J., Moose, S. P., Parsons, R. L., & Schmidt, R. J. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America*, 94(14), 7685–7690. <https://doi.org/10.1073/pnas.94.14.7685>

