Characterization of diverse carrot (*Daucus carota* L) genotypes using long noncoding RNA-based microsatellite markers

Abstract

In the pursuit of enhancing carrot breeding programs, our study focused on the experimental validation of selected long non-coding RNAs (lncRNAs) using end-point polymerase chain reaction (PCR). From an initial set, five lncRNAs were chosen for further experimental validation based on their bit score and the presence of simple sequence repeat (SSR) motifs. Our novel approach involved the development of 27 imperfect SSR markers and five perfect SSR markers specifically targeting these lncRNAs. The integration of SSR markers with lncRNA studies is unprecedented in carrot breeding, opening new avenues for genetic improvement. Among the five lncRNA-derived SSR (lncSSR) markers developed, two exhibited polymorphisms across 27 diverse carrot genotypes. In conclusion, our study highlights the innovative use of lnc-SSR markers in carrot breeding, demonstrating their utility in genetic diversity studies and their potential role in uncovering the functional importance of lncRNAs in crop phenotypes. The findings present significant implications for future carrot breeding strategies, providing a foundation for integrating molecular markers and lncRNA research in carrot and other related horticultural crop improvement programs.

Keywords: Carrot; Long non-coding RNA; Polymerase Chain Reaction; Simple Sequence Repeat; Polymorphism

Introduction

Long non-coding RNAs (lncRNAs), which exceed 200 nucleotides in length and do not code for proteins, play a crucial role in the regulation of gene expression at multiple levels. To classify a transcript as "noncoding," it is assessed based on the absence of an extensive open reading frame (typically over 100 codons) and the lack of codon conservation (Morris and Mattick, 2014). In the nucleus, lncRNAs regulate gene expression through transcriptional and epigenetic mechanisms, including the modulation of chromatin compaction and the establishment of chromosomal connections (Ma et al., 2022). They also influence gene expression by regulating the loading of histone or chromatin modifiers to chromatin. Additionally, lncRNAs can directly affect transcription through the formation of R-loops, interference with RNA polymerase machinery, and transcription of the lncRNA locus. In the cytoplasm, a subset of lncRNAs is involved in controlling mRNA turnover and translation, highlighting their multifaceted roles in gene expression regulation (Waititu et al., 2020).

Plant lncRNAs, transcribed by RNA polymerases PolI,PolII, and PolV, originate from various genomic regions, including enhancers, promoters, gene introns, pseudogenes, and as antisense transcripts to other genes (Axtella, 2013; Dekeba, 2021). These transcripts can be classified into different types based on their genomic position relative to protein-coding genes, such as natural antisense transcripts (lncNATs), intronic lncRNAs, intergenic lncRNAs (lincRNAs), and sense lncRNAs. lncNATs align with or diverge from sense strand transcripts, potentially modulating gene regulation, while intronic lncRNAs are transcribed from within introns of protein-coding genes without exon overlap. Sense lncRNAs share promoters with protein-coding genes and transcribe from regions overlapping exons (Ma et al., 2022; Meng et al., 2021). Recent studies have revealed the involvement of lncRNAs in various biological processes, including flowering, reproduction, photomorphogenesis, vernalization, organ development, cell cycle control, and responses to biotic and abiotic stresses. The experimental validation of lncRNAs is essential to elucidate their interactions with transcriptional and translational components, providing a strong rationale for exploring their regulatory roles in different metabolic pathways in plants.

Carrot (*Daucus carota* L.) is a biennial herb with a chromosome number of 2n = 2x = 18 with a genome size of 473Mb and belongs to the Apiaceae family (Iorizzo et al., 2016).Fewer studies are available on lncRNAs in carrots compared to model organisms like Arabidopsis despite the availability of high-quality genome sequences in carrots as large numbers of lncRNAs have been found in nature (Rai et al., 2019).

Material and methods

Plant material

The present study was conducted at the Department of Biotechnology and Crop Improvement, University of Horticultural Sciences, Bagalkot.Total of 27 genotypes carrot (*Daucus carota* L.) germplasm lines were grown, inclusive of Asiatic/Eastern and European/Western were used. The detailed names and colour description of the 27 carrot genotypes are listed in table 1. The vegetative phase carrot root evaluation experiment was conducted at sector 1 farm (16°12'N, 75°12'45 E), University of Horticultural Sciences (UHS), Bagalkot, Karnataka, India during 2022-2023. It is part of a semi-arid tropical region with a mild climate, with typical temperatures between 23 °C and 35 °C and receives an average rainfall of 318 mm.

Primer designing for lncRNAs

The lncRNAs were identified in the Alnc database and Primers were designed by searching microsatellite regions in the lncRNA transcripts. Sequences flanking the microsatellite regions in lncRNA were targeted for primer design using the Krait tool (Du et al., 2018) by targeting the microsatellite flanking regions. The primer pairs were designed based on standard criteria, such as predicted melting temperature of 50-60^{0}C and Guanine-Cytosine (GC) content of 45-55 percent for designing the primer pairs to obtain PCR amplicon length of 100-200 bp (Table 2).

Genomic DNA extraction and amplification of lncSSR markers

The genomic DNA was extracted from carrot leaves using CTAB extraction method (Doyle *et al*., 1991). DNA amplification was carried out using lncRNA-SSR primers for 27 genotypes of carrot to check their utility and functional validation. Amplification of DNA for the respective target lncRNA was performed using polymerase chain reaction (PCR) in a 10 µl reaction mixture for lncRNA validationeach containing 2 µl template DNA (50 ng/µl), 0.5 µl forward primers (5 µmol/l), 0.5 µl reverse primers (5 µmol/l), 2 µl ddH2O and 5 µl 2 × Taq PCR Master mix. The PCR program was as follows: 1 min at 95 °C, 20 s denaturing at 95 °C, 45 s annealing at 47-53 °C (Annealing temperature was optimised for each primer set using gradient PCR before actual amplification) and 30 s elongation at 72 °C followed by a final step at 72 °C for 2 min. The details of primer sequences used for amplification of lncRNA is presented in the table 3. After the PCR, fractionation was done using agarose gel electrophoresis to assess the amplification of target genes followed by EtBr staining. 3.5 % agarose gel was used for separation of lnc-SSR markersin 1 × TBE buffer at initial 85V for 15 minutes followed by 120 V for 1 h. Subsequently, the banding pattern was visualized using a gel documentation system and clear images of the bands were captured and scoring was performed.The clear and unambiguous bands of all the polymorphic lncSSR markers were scored for 27 carrot genotypes and calculated manually.

Results and Discussion

Identification of SSR regions in lncRNA and marker development

The identified sets of transcripts from the Alnc database were used to search for simple sequence repeats (SSR) in the Krait tool. 32 sequences were considered to contain microsatellites or SSR motifs. The identified SSRs have mono, di, tri, tetra, and hexa nucleotide repeat motifs. Among 32, perfect SSRs were five (Table 2a) and the remaining 27 were imperfect SSRs (Table 2b). From

the perfect SSRs- two SSRs contain mono- a nucleotide repeat motif, two SSRs contain di- a nucleotide repeats motif and the remaining one has hexa- nucleotide repeat motif. Among the imperfect SSRs- five SSRs contain mono- nucleotide repeat motif, eleven SSRs contain di- a nucleotide repeat motif, six SSRs contain tri- nucleotide repeat motif, four SSRs contain quad-nucleotide repeat motif and the remaining one has hexa- nucleotide repeat motif. One perfect SSR (STRG.55.1) was chosen for validation as it consists of a hexa-nucleotide repeat motif. Four among the imperfect SSRs, STRG.25.1 (di-nucleotide repeat motifs), STRG.62.1 and STRG.70.1 (tri-nucleotide repeat motifs), and STRG.91.1 (tetra nucleotide repeat motifs) were chosen for validation because of the presence of a greater number of repeats and higher chances of polymorphism.

Validation of lncRNA by PCR

A total of five SSR markers were used for the validation of 27 genotypes which yielded clear and scoreable bands. Among the five SSR markers considered for validation *Dc*Lnc91, *Dc*Lnc25 and *Dc*Lnc70 showed monomorphic bands of 140, 100, and 140bp respectively. While markers*Dc*LNC55 and *Dc*LNC62 showed polymorphism with band sizes ranging from 100-125 bp and 200-480 bp respectively (Table 4) and gel profile (Figure 1) indicating their suitability in the carrot genetic diversity assessment.For validation of lncRNA, 5 lnc-SSR markers were used.Among them, 3 lnc-SSR markers showed monomorphic bands, while 2 markers showed polymorphism among 27 diverse carrot genotypes. Polymorphic SSR markers can reveal genetic diversity or variations in lncRNA sequences among different genotypes and would be further useful for mapping and understanding its role in phenotypic expression of economic traits in carrots.

Genotypic scores for lnc-SSR markers

For lnc-SSR allele sizing, the genotypes were analyzed by comparing the sizes of the amplified bands against a 100bp reference ladder. Each genotype's band size was measured and recorded, ensuring accurate identification of allele variations. The use of the 100bp ladder provides a reliable standard for consistent and reproducible results across different samples.The observation of monomorphic bands in 3 and-SSR markers and polymorphic bands in two markers suggests potential genetic variation.

Conclusion

The identification of polymorphic bands in two out of five lncRNA SSR markers used for validation indicates genetic variation in the genotypes. This study underscores the pivotal role of long non-coding RNAs (lncRNAs) in the genetic and phenotypic diversity of carrot genotypes. By developing novel lncRNA-derived SSR markers, we have provided useful tools for assessing genetic variation in carrots. The polymorphism observed in these markers among diverse carrot genotypes highlights their potential application in carrot breeding programs. Our findings open new avenues for integrating molecular markers and lncRNA research in crop improvement strategies.

FUTURE SCOPE

Generating genome-wide lncRNAs and utilizing the polymorphic markers in future analyses will enhance insights into the genetic diversity of regulatory roles of these non-coding sequences. They help breeders to understand the functional roles and regulation of important genes influencing carrot morphology and other phenotypic characteristics. This knowledge can be leveraged to develop new carrot varieties with desirable traits, enhancing breeding efficiency and outcomes.

REFERENCES

Morris K. V. and Mattick J. S., 2014, The rise of regulatory RNA. Nat. Rev. Genet., 15(6): 423-437.

Ma X., Zhao F. and Zhou B., 2022, The characters of non-coding RNAs and their biological roles in plant development and abiotic stress response. Int. J. Mol. Sci., 23(8):4124.

Waititu J. K., Zhang C., Liu J. and Wang H., 2020, Plant non-coding RNAs: Origin, biogenesis, mode of action and their roles in abiotic stress. Int. J. Mol. Sci. 21(21): 8401.

Axtell M .J., 2013, Classification and comparison of small RNAs from plants. Annu. Rev. Plant Biol., 64: 137-159.

Dekeba D. G., 2021, Advances on the application of non-coding RNA in crop improvement. Afr. J. Biotechnol., 20(11): 440-450.

Meng X., Li A., Yu B. and Li S., 2021, Interplay between miRNAs and lncRNAs: Mode of action and biological roles in plant development and stress adaptation. Comput. Struct. Biotechnol. J., 19: 2567-2574.

Iorizzo M., Ellison S., Senalik D., Zeng P., Satapoomin P., Huang J., Bowman M., Iovene M., Sanseverino W., Cavagnaro P. and Yildiz M., 2016, A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat. Genet, 48(6):657-666.

Rai M. I., Alam M., Lightfoot D. A., Gurha P. and Afzal A. J., 2019, Classification and experimental identification of plant long non-coding RNAs. Genomics., 111(5): 997-1005.

Du L., Zhang C., Liu Q., Zhang X. and Yue B., 2018, Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics, 34(4):681-683.

Doyle, J., 1991. DNA protocols for plants. In Molecular techniques in taxonomy (pp. 283-293). Berlin, Heidelberg: Springer Berlin Heidelberg.

Table 1. List of carrot (*Daucus carota* L.) genotypes used for evaluation of genetic variability

Sl. No.	Genotypes	Description
1.	UHSBC-34-2	Orange
2.	UHSBC-17	Orange
3.	UHSBC-31-2	Orange
4.	UHSBC-31-1	Orange
5.	UHSBC-32-2	Orange
6.	UHSBC-65	Released variety, Red colour (Pusa Rudhira)
7.	UHSBC146	Orange (Improved-population-2)
8.	UHSBC-23-1-1	Orange-red
9.	UHSBC-66	Black, released (Pusa Asita)
10.	UHSBC-31-4	Orange
11.	UHSBC-34-2-1	Orange
12.	UHSBC-100	Dark Orange (Western)
13.	UHSBC-46-1-1	Orange
14.	UHSBC-151	Reddish orange
15.	UHSBC-23-1	Orange
16.	UHSBC-150	Reddish orange
17.	UHSBC-1	Pale orange
18.	UHSBC-152	Orange
19.	UHSBC-160	Orange
20.	UHSBC-17-2	Orange
21.	UHSBC-59-66	White
22.	UHSBC-59	White
23.	UHSBC-17-1	Orange
24.	UHSBC-117	Orange
25.	UHSBC-150-1	Orange

Table 2a. List of lncRNA perfect SSRs designed

Sl. No.	Sequence_ID	Entry	Product	Forward	Tm1	GC1	Stability1	Reverse	Tm2	GC2	Stability2
1	STRG.36.1	1	122	ACAAAGCTCCCCTCCC	54.77	62.5	4.3	CGGTGGATAACAGTGTGC	55.8	55.56	4.57
2	STRG.45.1	1	114	CCTTAATTCTATGATAAAGGCTGG	55.12	37.5	4.85	ACATGCAACCCCTAAAAGc	56.07	47.37	3.51
3	STRG.48.1	1	134	CCGGCCATGACCAAGTCC	60.44	66.67	3.85	GAGGGGGAGAGAGAGCGG	60.52	72.22	5.52
4	STRG.55.1	1	122	TTTCTTGGGTAAAGCAAGG	53.01	42.11	3.61	GACTTTCTTAAAGGCGATACC	54.41	42.86	2.73
5	STRG.70.1	1	145	AAAAGGGGAAAGTGCGGC	58.54	55.56	6.53	CAATGACCATTTTTATCAAACCCC	57.02	37.5	4.95

Table 2b. List of lncRNA imperfect SSRs

Sl. No.	Sequence ID	Forward	Tm1	GC1	Stability1	Reverse	Tm2	GC2	Stability2
1	STRG.8.1	CCTCTGCTCTGCTAAGCTCC	59.9	60	4.7	ATCGGAATCTGAGCGGCG	59.97	61.11	6.46
2	STRG.8.1	CCTCTGCTCTGCTAAGCTCC	59.9	60	4.7	ATCGGAATCTGAGCGGCG	59.97	61.11	6.46
3	STRG.24.1	AAACAACAATGCCTGGGC	56.48	50	5.36	CTCCAGGAGAGTTTTAGTAGG	54.49	47.62	3.18
4	STRG.25.1	TGCTGTTTGCTTTCCCGG	58.56	55.56	5.73	GCGTGTTTATTCCAAAATGGCC	59.33	45.45	5.36
5	STRG.27.1	TCCAACCCAGCGATTCGG	59.73	61.11	4.3	TCTCCGATGAACACCGCG	59.82	61.11	6.46
6	STRG.34.1	CACGGGCGTTTGAATGGC	60.13	61.11	4.4	TCACACGTTCCGAAGAGCC	60.01	57.89	4.7
7	STRG.36.1	ACAAAGCTCCCCTCCC	54.77	62.5	4.3	CGGTGGATAACAGTGTGC	55.8	55.56	4.57
8	STRG.40.1	CAATCCTATTATTCAACCCCC	53.68	42.86	5.4	TTCTCGGTGTTGTGTGC	54.94	52.94	4.57
9	STRG.45.1	CCTTAATTCTATGATAAAGGCTGG	55.12	37.5	4.85	ACATGCAACCCCTAAAAGC	56.07	47.37	3.51
10	STRG.48.1	CCGGCCATGACCAAGTCC	60.44	66.67	3.85	GAGGGGGAGAGAGAGCGG	60.52	72.22	5.52
11	STRG.48.1	CCGGCCATGACCAAGTCC	60.44	66.67	3.85	TCAAGAGTGGGAGAGAGAGGG	59.99	57.14	4.3
12	STRG.48.1	TTATCTCTCTCCCTCTCC	50.64	50	3.71	AGAGAGAGAGAGTGGG	49.42	56.25	4.61
13	STRG.49.1	GCACCAAGCAAATTTTCGG	56.29	47.37	4.3	GGCGGATCTAGGAAAAGGC	57.99	57.89	4.35
14	STRG.50.1	GTGAGGTCTGCGGCGG	60.15	75	6.13	ACCCTCGACAAATTCAAATAGGC	59.31	43.48	3.93
15	STRG.51.1	TGTCAAGGCAGGACAGAAGC	60.25	55	3.86	GTAAAGGGCAAGGCAGGC	58.71	61.11	4.85
16	STRG.53.1	TGAATAGGACTGCGAGAGAAAGG	59.87	47.83	3.11	AGTAAGGAGGGGCGTACG	58.07	61.11	3.67
17	STRG.55.1	TGATTTTCTTGGGTAAAGCAAGG	57.53	39.13	3.61	AGCTTAAGACACATCCAATCCC	58.11	45.45	3.85
18	STRG.62.1	CTCCTTTATAATTTAACAGGTGGG	55.23	37.5	4.61	CCTCCAAAGTCCAAAGTGC	56.42	52.63	4.4
19	STRG.62.1	TCTTCCTTTGATTATCCCA	57.88	43.48	4	TGTGCAAACCTCCAAA	58.6	50	3.85

Sl. No.	Sequence ID	Forward	Tm1	GC1	Stability1	Reverse	Tm2	GC2	Stability2
		CAGG				GTCC			
20	STRG.66.1	CCCTCTTTCTCAGCGCC	57.3	64.71	6.53	CGACGATCTCGAAATCATGG	56.53	50	3.66
21	STRG.70.1	AAAAGGGGAAAGTGCGGC	58.54	55.56	6.53	CAATGACCATTTTTATCAAACCCC	57.02	37.5	4.95
22	STRG.70.1	AATTGTGGGGTTTGATAAAAATGG	56.61	33.33	3.16	AATTTTGCCCCCACCACC	58.1	55.56	4.61
23	STRG.75.1	TCACTACACTAGTCTCTCTCTCC	57.65	47.83	3.71	AGGTGTATGCAGAGAGGCC	58.78	57.89	5.19
24	STRG.78.1	CCAACAACCCAGCAGCCC	61.31	66.67	5.19	GGTGTGTCTCCCCAGTAATGG	60.07	57.14	3.16
25	STRG.84.1	AGCTCGCTCACTTCTGGC	59.74	61.11	4.85	GCGACAATGATTTCTCCGGC	59.97	55	6.13
26	STRG.91.1	GGGAGGGAGGGAAGATCC	57.68	66.67	3.36	ATTATCACAATGCTCTTTTCTTCC	55.74	33.33	3.46
27	STRG.91.1	GTGATAATGTAAGTAAGTAGCTGC	55.46	37.5	5.25	TTTGTTTCCCCTCCTTGC	54.99	50	4.01

Table 3. The details of lnc-SSR markers used for validation in endpoint PCR

Sl. No.	Primer ID	Forward primer sequence (5`-3`)	Reverse primer sequence (5`-3`)
1.	*Dc*LNC55	TTTCTTGGGTAAAGCAAGG	GACTTTCTTAAAGGCGATACC
2.	*Dc*LNC62	CTCCTTTATAATTTAACAGGTGGG	CCTCCAAAGTCCAAAGTGC
3.	*Dc*LNC91	GTGATAATGTAAGTAAGTAGCTGC	TTTGTTTCCCCTCCTTGC
4.	*Dc*LNC25	TGCTGTTTGCTTTCCCGG	GCGTGTTTATTCCAAAATGGCC
5.	*Dc*LNC70	AAAAGGGGAAAGTGCGGC	CAATGACCATTTTTATCAAACCCC

Table 4. List of the lnc-SSRs amplification data in carrot (*Daucus carota* L.) genotypes

Sl. No.	Primer ID	Annealing temperature (℃)	Observed amplicons size (bp)
1.	*Dc*LNC55	47	100-125
2.	*Dc*LNC62	48	200-480
3.	*Dc*LNC91	48	140
4.	*Dc*LNC25	53	100
5.	*Dc*LNC70	50	140