Journal of Medicinal Materials, 2026, Vol. 31, No. 1+2 (pp. 154 - 160)

COMPARATIVE TRANSCRIPTOME PROFILING OF *MOMORDICA COCHINCHINENSIS* (LOUR.) SPRENG. ARIL USING OXFORD NANOPORE AND ILLUMINA SEQUENCING TECHNOLOGIES

***Tran Duc Trung*[1], *Nguyen Hoang*[1], *Le Thi Tu Linh*[1], *Nguyen Thi Thu Hang*[1],**
Pham Thanh Huyen*[1], *Truong Nhat My*[2], *Nguyen Thi Lan Hoa*[1,]
[1]*National Institute of Medicinal Materials (NIMM), Hanoi 11018, Vietnam;*
[2]*Vietnamese-German Center for Medical Research, 108 Institute of Clinical Medical and Pharmaceutical Science, Hanoi, Vietnam*
*Corresponding author: nguyen.lanhoa@gmail.com

Received November 20th, 2025 Accepted January 19th, 2026

Summary

Gac (*Momordica cochinchinensis* (Lour.) Spreng.) is a tropical fruit recognized for its exceptionally high carotenoid content, particularly lycopene and β-carotene. Although Gac possesses significant nutritional and pharmacological value, genomic and transcriptomic resources for this species are limited. Previous studies utilizing Illumina RNA-Seq established the first *de novo* transcriptome of Gac aril, identifying over 81,000 unigenes and key enzymes involved in carotenoid biosynthesis. In this study, we generated a transcriptome of Gac aril using Oxford Nanopore Technologies (ONT) direct RNA sequencing and compared its assembly and annotation with the published Illumina dataset. ONT sequencing produced over 13 million long reads, resulting in 52,755 unigenes with an average length of 777 bp and an N50 of 922 bp, which are substantially longer than those generated using Illumina sequencing (average 388 bp, N50 450 bp). Functional annotation identified 33,235 unigenes with significant matches, including full-length carotenoid-related unigenes, e.g., phytoene synthase (*PSY*), phytoene desaturase (*PDS*), zeta-carotene desaturase (*ZDS*), zeta-carotene isomerase (*Z-ISO*), prolycopene isomerase (*CRISTO*), lycopene ß-cyclase (*LCYB*), and lycopene ε-cyclase (*LCYE*). Long-read ONT sequencing enabled the recovery of full-length transcripts, reducing assembly fragmentation and redundancy and yielding a more compact, biologically representative unigene set. While Illumina technology remains superior in sequencing accuracy and quantitative profiling, ONT offers enhanced structural resolution, isoform discovery, and the potential to detect RNA modifications. Recent advances in ONT chemistry, flow cell design, and bioinformatic pipelines are rapidly improving accuracy and throughput, establishing ONT as a credible stand-alone approach for transcriptome analysis in non-model plants. These findings demonstrate that ONT sequencing can reveal novel regulatory features of carotenoid biosynthesis in Gac, providing valuable resources for functional genomics and metabolic engineering in carotenoid-rich crops.

Keywords: *Momordica cochinchinensis; Gac aril; Oxford Nanopore Technologies; RNA-seq; Comparative transcriptome*

1. Introduction

Carotenoids are essential pigments in plants that play crucial roles in photosynthesis, photoprotection, and signaling. In addition to their physiological functions, carotenoids possess significant nutritional and medicinal value for humans, serving as antioxidants and precursors of vitamin A. Among carotenoid-rich species, Gac (*Momordica cochinchinensis* (Lour.) Spreng.), a perennial vine native to Southeast Asia, is notable for its exceptionally high concentrations of lycopene and *β*-carotene in the aril layer surrounding its seeds [1],[2]. These compounds are responsible for the fruit's vibrant red color and have been used in traditional medicine for health benefits, including reducing the risk of chronic diseases and improving eye health. As a result, Gac has emerged as a promising candidate for functional genomics and metabolic engineering to enhance carotenoid biosynthesis [3],[4].

Comprehensive transcriptome profiling is essential for elucidating the molecular mechanisms underlying carotenoid accumulation in plants. This approach enables the characterization of gene expression patterns, regulatory networks, and alternative splicing events that influence metabolic pathways. For non-model species lacking a reference genome, such as Gac, *de novo* transcriptome assembly is critical for gene discovery and functional annotation. These analyses yield insights into carotenoid biosynthetic pathways, identify candidate genes for breeding, and support the development of molecular markers for crop improvement [5]. Next-generation sequencing (NGS), particularly Illumina technology, has been the standard for RNA-seq analysis due to its high accuracy, throughput, and robust bioinformatics support. An early study by Hyun *et al.* [6] utilized Illumina short-read sequencing to characterize Gac transcriptomes, generating over 51 million high-quality reads from aril tissue and assembling 81,404 unigenes with an average length of 388 bp. This foundational research identified key enzymes in the carotenoid biosynthetic pathway, including phytoene synthase (*PSY*), phytoene desaturase (*PDS*), ζ-carotene desaturase (*ZDS*), carotenoid isomerase, and lycopene ε-cyclase (*LCYE*). Although Illumina sequencing provides reliable quantitative expression data, the short read length limits isoform resolution and impedes the detection of structural variants and RNA modifications.

Oxford Nanopore Technologies (ONT) has emerged as a complementary sequencing platform, providing long-read capabilities that facilitate the capture of full-length transcripts, the resolution of complex isoforms, and the direct detection of RNA modifications [7],[8]. ONT has been successfully applied to transcriptomic studies in plants such as *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays*, revealing full-length transcripts, novel isoforms, and alternative splicing events that are frequently overlooked by short-read sequencing [9],[10],[11]. Although ONT exhibits lower base-calling accuracy than Illumina, ongoing improvements in nanopore chemistry and bioinformatics pipelines have substantially increased its reliability [12]. The capacity to generate comprehensive transcript structures makes ONT particularly suitable for investigating secondary metabolite pathways, where isoform diversity and regulatory complexity are important.

Here, we presented a comparative analysis of Gac aril transcriptomes generated using the ONT and Illumina sequencing platforms. By integrating long-read ONT data with an established Illumina dataset, we assessed the strengths and limitations of each platform in reconstructing carotenoid biosynthetic pathways. This study presented the first comprehensive comparison of ONT and Illumina transcriptomic datasets in *M. cochinchinensis.* Our findings elucidated the distinct advantages of the long-read ONT platform in improving transcriptome assembly quality, thereby advancing the precision of functional genomics and facilitating targeted metabolic engineering in carotenoid-enriched crops.

2. Materials and methods

2.1. Plant materials

Aril tissues of Gac fruits (elite line GM10 - preserved at Gac Conservation Garden of Research Center of Medicinal Plants - National Institute of Medicinal Materials) were collected at mid-ripening stage (45 days after flowering, DAF), consistent with the developmental stage used in the Illumina study by Hyun *et al.* [6]. Samples were immediately frozen in liquid nitrogen and stored at -80°C.

2.2. RNA extraction

Total RNA was extracted from aril samples using RNAzol RT (Sigma-Aldrich, USA) supplemented with DNase I treatment to remove genomic DNA contamination. The polyA$^+$ mRNAs were enriched using Dynabeads™ mRNA Purification Kit (Thermo Fisher Scientific, USA) according to the manufacturer’s instructions. RNA integrity and quantity were assessed using agarose gel electrophoresis and Qubit fluorometric quantification, respectively.

2.3. Oxford Nanopore Technologies (ONT) sequencing

For library construction, 100 ng polyA$^+$ mRNA was used as the input for the ONT PCR-cDNA Barcoding Kit (SQK-PCB109, Oxford Nanopore Technologies, UK) in accordance with the manufacturer’s instructions. cDNA libraries were enriched using LongAmp Taq (New England Biolabs, USA) with thermocycling conditions as follows: initial denaturation at 95 °C for 30 s; 14 cycles of 95˚C for 15s, 62˚C for 15s, and 65˚C for 4 min; and a final extension at 65˚C for 10 min. Subsequently, library purification was performed using a 0.6× (v/v) ratio of AMPure XP magnetic beads (Beckman Coulter, USA) to remove residual contaminants and enrich the targeted cDNA for the downstream sequencing. Purified libraries were loaded onto an R9.4.1 flow cell and sequenced on the MinION Mk1B platform for 72 h. Basecalling and demultiplexing were performed using Dorado v0.7.2 (https://github.com/nanoporetech/dorado) in super-accuracy mode (“dna_r9.4.1_e8_sup@v3.6”).

2.4. ONT-based Gac aril transcriptome assembly

ONT sequencing reads were selected by Pychopper (https://github.com/nanoporetech/pychopper) using default parameters. Then, poly (A) tails were removed with Cutadapt using the *--poly-A* flag [13]. Reads retained for downstream analyses met stringent criteria, requiring a minimum quality score of Q7 and a length threshold of at least 100 bp. High-quality, filtered reads were then assembled *de novo* with RNA-Bloom2, utilizing recommended parameters for

Oxford Nanopore data (*-fpr 0.005 -overlap 200 -length 150 -lrop 0.7 -p 0.7 -lrrd 3 -e 3*) [14]. Assembly completeness was evaluated using BUSCO v.6.0.0, benchmarking against both Viridiplantae (v10) and Eudicots (v10) lineage datasets. Functional annotation of assembled transcripts was performed using BLASTx (v2.16.0) against the NR, Swiss-Prot, KEGG, Clusters of Orthologous Groups (COG), and Gene Ontology (GO) databases, with a significance threshold of 1e-5.

2.5. Comparative transcriptome profiling analysis

To rigorously assess the capabilities in Gac aril transcriptome reconstruction, the ONT-based assembly was benchmarked against the established Illumina-based dataset described by Hyun *et al.* [6]. Comparative metrics included total unigene count, mean transcript length, and contig/unigene N50 values, providing a quantitative framework for evaluating assembly quality. Functional annotation was further integrated by mapping assembled sequences to key databases, including NR, Swiss-Prot, KEGG, COG, and GO, to assess annotation depth comprehensively. Particular emphasis was placed on the identification and characterization of carotenoid biosynthetic pathway genes, such as *PSY, PDS, ZDS,* carotenoid isomerase, and *LCYE*, to elucidate the platforms' respective strengths in capturing full-length gene models and isoform diversity.

3. Results

3.1. ONT sequencing and de novo assembly of Gac aril transcriptome

Sequencing of cDNA from *M. cochinchinensis* aril using Oxford Nanopore Technologies (ONT) generated over 13 million reads, totaling 4.8 Gbp. Read lengths ranged from 50 bp to 50 kb, with mean and N50 values of 394 bp and 682 bp, respectively. The estimated error rate was 1.9%, and more than 89% of reads had a Qscore above 10. After full-length selection and quality filtering ($Q \geq 7$, length ≥ 100 bp), 7.6 million reads (58.9%) with an N90 of 790 bp were retained for assembly. *De novo* assembly with RNA-Bloom2, followed by redundancy reduction with CD-HIT, yielded 52,755 unigenes with an N50 of 922 bp. The assembled transcriptome totaled 41 Mb, highlighting the ability of ONT to capture full-length transcripts and represent isoform diversity accurately [7],[8]. Most transcript lengths were below 1,500 bp (93.5%), while only 1.1% exceeded 2,000 bp (Fig. 1).

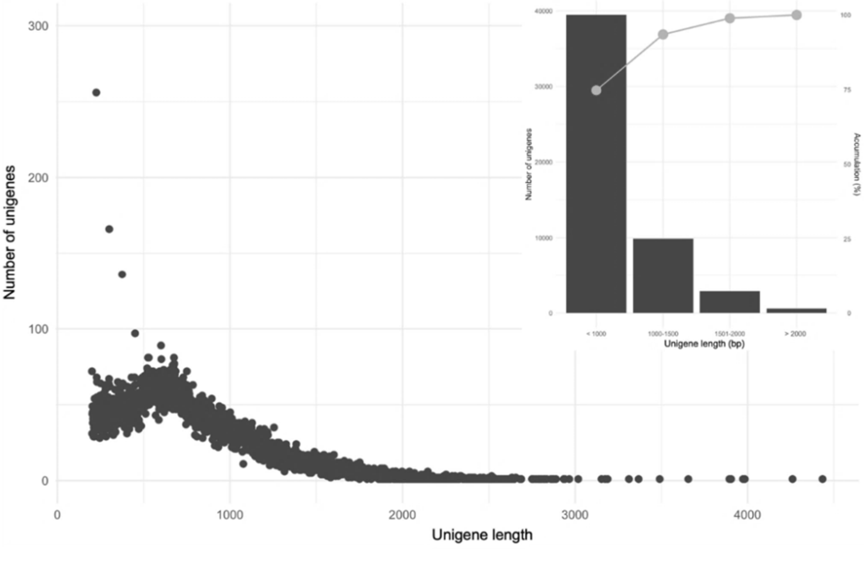

Fig. 1. Unigene length distribution of ONT-based Gac aril transcriptome at 45 DAF

3.2. Comparison between ONT-based and Illumina-based Gac aril transcriptomes

The Illumina-based transcriptome reported by Hyun *et al.* [6] generated 51.4 million short reads totaling 4.63 Gbp, which were assembled into 81,404 unigenes with an average length of 388 bp and an N50 of 450 bp. In comparison, ONT assemblies produced fewer unigenes but with significantly longer average lengths and higher N50 values, resulting in more complete transcript structures (Table 1).

Table 1. Comparative overview of sequencing approaches and Gac aril transcriptomes at 45 DAF

	Illumina-based approach [6]	**ONT-based approach (this study)**
Total raw reads	51,466,670	13,046,721
Total raw nucleotide (nt)	4,632,000,300	4,486,683,391
GC percentage (%)	45.38	41.86
Transcriptome - Total assembled unigenes	81,404	52,755
Transcriptome - Average length of unigenes (nt)	388	777.7
Transcriptome - N50 (nt)	450	922
Transcriptome - Total assembly size (Mbp)	31.579	41.026

3.3. Functional annotation of ONT-based transcriptome

Transcriptome completeness was evaluated using BUSCO (Fig. 2). Only 40% of the unigenes were classified as complete, a relatively low proportion that is expected given the nature of aril tissue, which lacks the transcriptional activity typical of photosynthetically and developmentally active tissues. Additionally, the length distribution of the newly assembled transcriptome reveals numerous fragmented genes, with most transcripts shorter than 1,500 base pairs. Although the ONT technique is theoretically capable of capturing full-length transcripts, several biases, including 5’ truncation, reverse transcription bias, and PCR bias, can result in transcript fragmentation.

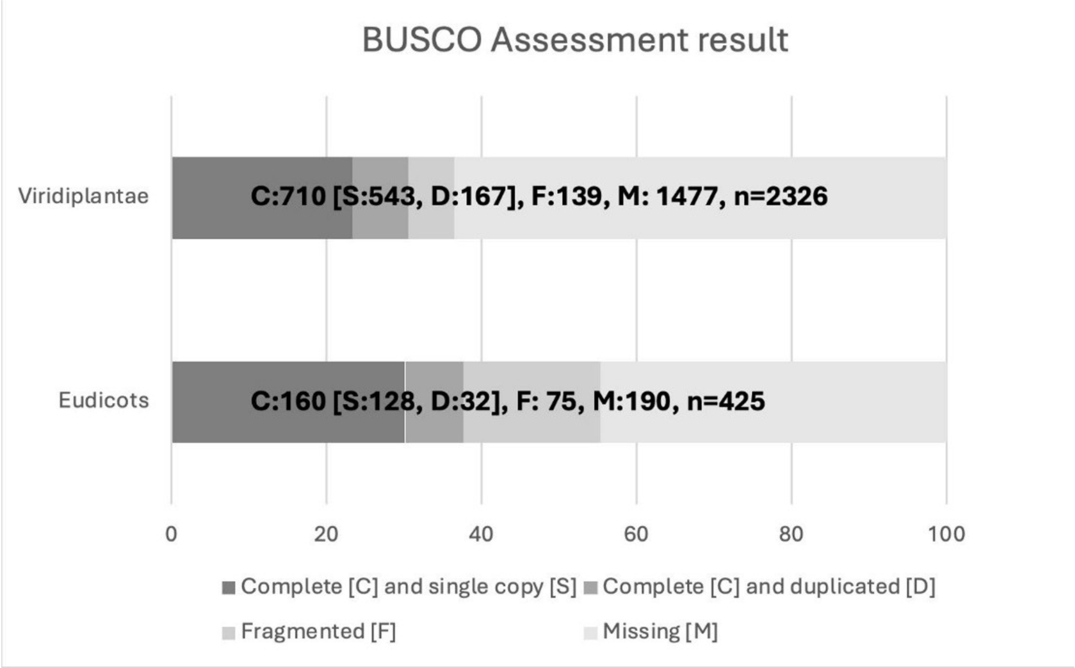

Fig. 2. BUSCO analysis of ONT-based Gac aril transcriptome at 45 DAF

Annotation of the ONT-based transcriptome against the NR, Swiss-Prot, KEGG, and COG databases identified 33,235 transcripts (63%) with significant matches (Table 2). This finding indicates that many genes of unknown function may contribute to the development of Gac aril, reflecting a complex biological process. The most frequent BLAST hits in the NR database were to *Momordica charantia*, followed by *Cucumis melo*, *Cucumis sativus*, and other Cucurbitaceae species. Of the 33,235 hits, 20,124 were assigned to COG classifications (Fig. 3). Because a single unigene can be assigned to multiple categories, the total number of categories may exceed the number of hits. Based on sequence homology, the largest group was “Poorly characterized” (4,428 unigenes), followed by “Post-translational modification, protein turnover, chaperones” (1,085 unigenes), and “Signal transduction mechanisms” (977 unigenes). Notably, 279 unigenes were classified under “Secondary metabolite biosynthesis, transport, and catalysis,” suggesting that these metabolic processes may play a key role in Gac aril development.

Table 2. Statistics of annotation results against the public databases

Public databases	Number of unigenes (this study)	Percentage of unigenes (this study)	Percentage of unigenes in the Illumina-based approach [6]
NR	32,093	60.83	43.38
Swiss-Prot	24,045	45.58	29.13
KEGG	21,345	40.46	21.01
COG	20,124	38.14	11.89
GO	27,378	51.89	25.08
All	33,235	63.00	

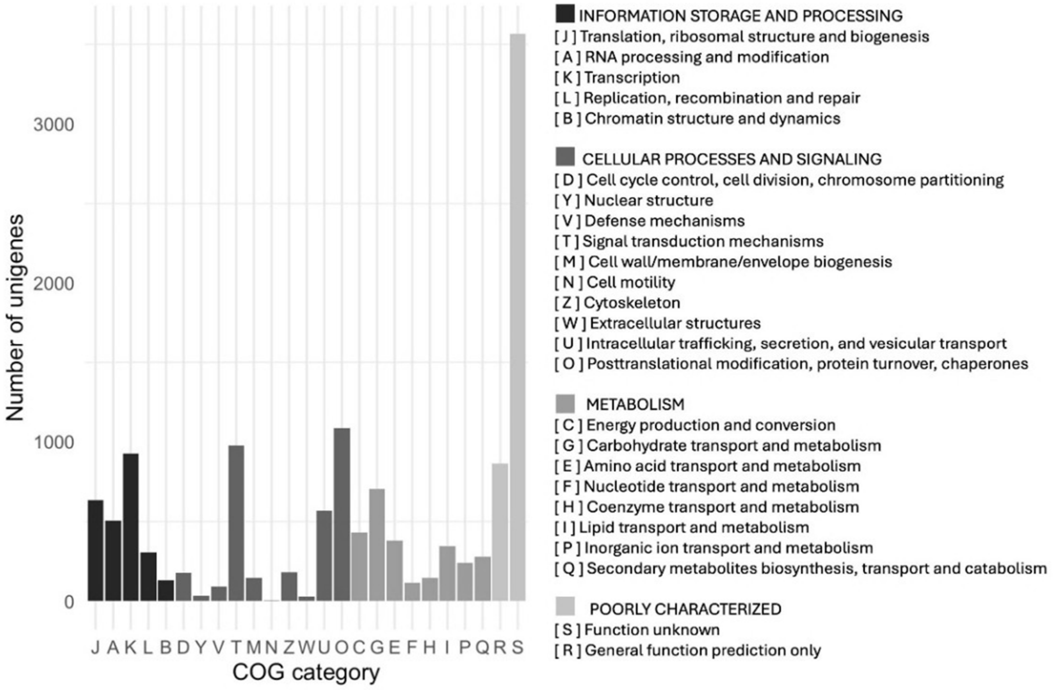

Fig. 3. COG assignment of ONT-based Gac aril transcriptome at 45 DAF

Comparison of the functional annotation results from this study with those reported by Hyun *et al.* [6] demonstrated a substantial improvement (Table 3). However, as the previous study was conducted in 2012, some aspects of the direct comparison might be irrelevant due to the differences in available data and methodologies. Despite this limitation, the broader annotation range of the newly assembled transcriptome underscores the significant potential of the ONT approach.

Functional annotation identified unigenes in the ONT-based transcriptome corresponding to carotenoid biosynthetic genes, including phytoene synthase (*PSY*), phytoene desaturase (*PDS*), zeta-carotene desaturase (*ZDS*), zeta-carotene isomerase (*Z-ISO*), prolycopene isomerase (*CRISTO*), lycopene ß-cyclase (*LCYB*), and lycopene ε-cyclase (*LCYE*). Compared with the study by Hyun *et al.* [6], the current analysis recovered all structural genes associated with carotenoid biosynthesis (Table 3). The application of long-read Nanopore sequencing facilitated the recovery of full-length transcripts, minimized assembly fragmentation and redundancy, and produced a more compact and biologically representative unigene set. In contrast, short-read Illumina assemblies often yield more unigenes due to fragmented contigs, partial transcripts, and redundant isoforms generated during *de novo* assembly [15]. This outcome is generally interpreted as an improvement in transcript completeness rather than a reduction in transcriptome diversity.

Table 3. Unigenes potentially related to carotenoid-related

Gene name	Gene abbreviation	Number of unigenes	
		ONT-based approach (this study)	**Illumina-based approach [6]**
Phytoene synthase	*PSY*	1	3
Phytoene desaturase	*PDS*	2	4
Zeta-carotene desaturase	*ZDS*	1	-
Zeta-carotene isomerase	*Z-ISO*	5	1
Prolycopene isomerase	*CRISTO*	1	3
Lycopene ß-cyclase	*LCYB*	1	-
Lycopene ε-cyclase	*LCYE*	1	1

4. Discussion

This study presents the to our knowledge, this is the first comparison between ONT and Illumina sequencing platforms for transcriptome profiling of *M. cochinchinensis* aril. Illumina RNA-Seq, as reported by Hyun *et al.* [6], generated over 81,000 unigenes with an average length of 388 bp and an N50 of 450 bp, resulting in a comprehensive catalog of expressed genes. In contrast, ONT sequencing produced approximately 52,000 unigenes, characterized by significantly longer average lengths (777 bp) and higher N50 values (922 bp). These findings highlight the trade-off between sequencing depth and read length, with Illumina providing greater coverage and ONT offering improved structural resolution. Further, as in the Illumina approach, the ONT direct RNA sequencing preferentially captures polyadenylated transcripts, introducing a known bias against non-polyadenylated RNAs [16]. This technical limitation is unlikely to affect the primary target and results of the presented findings, as the genes involved in carotenoid biosynthesis in Gac are protein-coding and therefore polyadenylated and well represented in our dataset. Nonetheless, it should be noted that non-polyadenylated RNAs, including potential regulatory elements, may be underrepresented and could be explored in future work using broader library preparation strategies.

The observed diversity of isoforms in carotenoid biosynthetic genes suggests the existence of novel regulatory mechanisms underlying the high carotenoid accumulation in Gac. Isoform-specific expression of *PSY* and *LCYE* may influence flux through the carotenoid pathway, thereby affecting lycopene and β-carotene concentrations. ONT's ability to detect RNA modifications further indicates the potential for epitranscriptomic regulation of carotenoid biosynthesis. These results suggest new opportunities for metabolic engineering, in which targeted interventions at the isoform or modification level could enhance carotenoid production in biofortification programs [17].

The long-read capability of ONT enabled the reconstruction of full-length transcripts, capturing isoform diversity and alternative splicing events not resolved in the Illumina dataset. For example, multiple isoforms of phytoene synthase (*PSY*) and zeta-carotene isomerase (*Z-ISO*) were identified, suggesting isoform-specific regulation of carotenoid biosynthesis. These structural insights are significant, as alternative splicing modulates plant metabolic pathways [5]. However, ONT sequencing has notable limitations. Its base-calling accuracy is lower than that of Illumina, especially in homopolymer regions and indel-rich sequences [12]. This limitation requires extensive error correction using tools such as RATTLE or a hybrid assembly approach with Illumina reads [18]. Furthermore, ONT throughput is lower, resulting in fewer total unigenes compared to the Illumina unigene catalog. Reduced sequencing depth may impede the detection of low-abundance transcripts, which often play critical roles in regulatory networks.

The application of ONT sequencing to non-model plants such as *M. cochinchinensis* demonstrates both its potential and its challenges. ONT provides superior structural resolution, facilitates isoform discovery, and enables epitranscriptomic analysis, but requires rigorous error correction, adequate sequencing depth, and robust quantification strategies. As ONT chemistry and basecalling algorithms improve, their utility in functional genomics is likely to increase, particularly for species with limited genomic resources. For non-model species like Gac, which lack reference genomes and comprehensive transcriptomic data, ONT's ability to generate full-length transcripts without assembly bias is especially advantageous. With ongoing improvements in sequencing kits and bioinformatic tools, ONT is emerging as a credible stand-alone approach for transcriptome analysis. This development suggests that ONT will not only complement Illumina but may also become a primary platform for functional genomics in under-studied crops, enabling the identification of novel isoforms, regulatory mechanisms, and RNA modifications that contribute to valuable metabolic traits such as carotenoid biosynthesis.

5. Conclusion

This study presented the first direct comparison between ONT and Illumina sequencing platforms for transcriptome profiling of *M. cochinchinensis* aril. Illumina RNA-Seq remains the benchmark for accuracy and quantitative profiling, while ONT offers distinct advantages in capturing full-length transcripts, isoform diversity, and RNA modifications. Despite current limitations in accuracy and throughput, rapid advances in ONT chemistry, sequencing kits, and bioinformatic pipelines are steadily improving performance. For non-model plants such as Gac, ONT represents a promising stand-alone approach to uncover transcript structures and regulatory mechanisms underlying carotenoid biosynthesis. As the technology matures, ONT is poised to become a credible option for functional genomics and metabolic engineering in under-studied crops.

Funding: *This work was supported by the Vietnam Ministry of Science and Technology (NVQG-2022/ĐT.09).*

References

1. Vuong L.T. (2000), Underutilized *β*-carotene–rich crops of Vietnam. *Food and Nutrition Bulletin*, 21(2), 173-181.

2. Do T. V. T., Fan L., Suhartini W., Girmatsion M. (2019), Gac (*Momordica cochinchinensis* Spreng) fruit: A functional food and medicinal resource. *Journal of Functional Foods*, 62, 103512.

3. Chuyen H. V., Nguyen M. H., Roach P. D., Golding J. B., Parks S. E. (2014), Gac fruit (*Momordica cochinchinensis* Spreng.): a rich source of bioactive compounds and its potential health benefits. *International Journal of Food Science & Technology*, 50(3), 567-577.

4. Mat S. A., El Enshasy H. A., Abd Rahim N., Chong X. N., Vadiveloo S. D., Ya'akob H., Daililin J. D., Chew S. T. D., Manas N. H. A., Gac fruit (*Momordica cochinchinensis* Spreng.): From nutritional value to food processing technology. *Biocatalysis and Agricultural Biotechnology*, 62, 103444.

5. Jackson D.J., Cerveau N., Posnien N. (2024), De novo assembly of transcriptomes and differential gene expression analysis using short-read data from emerging model organisms – a brief guide. *Frontiers in Zoology*, 21(1), 17.

6. Hyun T. K., Rim Y., Jang H. J., Kim C. H., Park J., Kumar R., Lee S., Kim C. B., Bhak J., Nguyen Q. B., Kim S. W., Lee Y. S., Kim J. Y. (2012), De novo transcriptome sequencing of *Momordica cochinchinensis* to identify genes involved in the carotenoid biosynthesis. *Plant Molecular Biology*, 79(4-5), 413-27.

7. Garalde D. R., Snell E. A., Jachimowicz D., Sipos B., Lloyd J. H., Bruce M., Lloyd J., Warland A., Pantic N., Admassu T., Ciccone J., Serra S., Keena J., Martin S., McNeil L., Wallace J., Jayasinghe L., Wright C., Blasco J., Sipos B., Young S., Juul S., Clarke J., Turner D. J. (2018), Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15(3), 201-206.

8. Workman E. R., Tang D. A., Tang S. P., Jain M., Tyson R. J., Razaghi R., Zuzarte C. P., Gilpatrick T., Payne A., Quick J, Sadowski N., Holmes N., Jesus G. J., Jones L. K., Soulette M. C., Snutch P. T., Loman N., Paten L., Loose M., Simpson T. J., Olsen H. H, Brooks N. A., Akeson M., Timp W. (2019), Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, 16(12), 1297-1305.

9. Yu F., Qi H., Gao L., Luo S., Njeri D. R., Ke Y., Wu P., Yang P. (2023), Identifying RNA modifications by direct RNA sequencing reveals complexity of epitranscriptomic dynamics in rice. Genomics, *Proteomics & Bioinformatics*, 21(4), 788-804.

10. Xuhui L., Weiwei C., Siqi L., Junteng F., Hang Z., Xiangbo Z., Yongwen Q. (2022), Full-length transcriptome analysis of maize root tips reveals the molecular mechanism of cold stress during the seedling stage. *BMC Plant Biology*, 22(1), 398.

11. Wang S., Shi Y., Zhou Y., Hu W., Liu F. (2024), Full-length transcriptome sequencing of *Arabidopsis* plants provided new insights into the autophagic regulation of photosynthesis. *Scientific Reports*, 14(1), 14588.

12. Soneson C., Yao Y., Bratus-Neuenschwander A., Patrignani A., Robinson M. D., Hussain S. (2019), A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications*, 10(1), 3359.

13. Martin M. (2011), Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10-12.

14. Nip K. M., Hafezqorani S., Gagalova K. K., Chiu R., Yang C., Warren R. L., Birol I. (2023), Reference-free assembly of long-read transcriptome sequencing data with RNA-Bloom2. *Nature Communications*, 14(1) 2940.

15. Fang Y., Chen G., Chen F., Hu E., Dong X., Li Z., He L., Sun Y., Qiu L., Xu H., Cai Z., Liu X. (2011), Accurate transcriptome assembly by Nanopore RNA sequencing reveals novel functional transcripts in hepatocellular carcinoma. *Cancer Science*, 112(9): p. 3555-3568.

16. Viscardi M. J., Arribere J. A. (2022), Poly(a) selection introduces bias and undue noise in direct RNA-sequencing, *BMC Genomics*, 23, 530.

17. Giuliano G., Tavazza R., Diretto G., Beyer P., Taylor M. A. (2008), Metabolic engineering of carotenoid biosynthesis in plants. *Trends in Biotechnology*, 26(3), 139-145.