Moringa Functional Genomics: Implications of Long Read Sequencing Technologies

11

J. Deepa, Rohit Kambale, A. Bharathi, M. Williams, N. Manikanda Boopathi, and M. Raveendran

Abstract

The caloric needs of the global population are solely dependent on only 30 crops including rice, wheat, and maize. It is important to diversify and ensure the global food supply by enhancing the wide adaptation of nutritionally important vegetable crops. Focusing on the underutilized and locally available vegetable crops would meet the need of eradicating malnutrition and food security for the local population and small-hold farmers. With the advent of long-read sequencing technologies, several orphan crops and crops with complex genetic nature are being explored and genome resources are made available for crop improvement. *Moringa oliefera* is a well-known medicinal and nutritive plant which is getting adequate attention in recent years. Functional analytical tools including genome, transcriptome, metabolome, and proteome were made available for moringa; besides diversity analysis, evolutionary and syntenic studies were also reported. With an objective of exploring the available resources of an important nutritive crop and evaluating the need to enhance the functional studies further, this chapter reviews the application of long-read sequencing and their application in moringa genomics for its wider utilization in pharma industries.

11.1 Prelude on Functional Genomics Tools and Applications in Moringa

11.1.1 Tools and Strategies Available to Explore Moringa Functional Genomics

Promoting the research of orphan crops by generating genome and transcriptome sequences will provide insights into the genes involved in important agronomic traits (Jamnadass et al. 2020). Moringa (*Moringa oleifera*) is an important vegetable crop in many developing countries due to its medicinal and nutritional properties. Moringa is a softwood tree that belongs to the family *Moringaceae,* and it is originated from the sub-Himalayan region of India (Ramachandran et al. 1980). The Moringa plant is gaining attention in the recent past due to its nutritional, stress-tolerant, economical, and medicinal values. Each part of Moringa is reported to possess medicinal value including leaf extract, seed, stem, flower, and root (Anwar et al. 2007; Sehgal et al. 2012; Al-Asmari et al. 2015). Genome and

J. Deepa · R. Kambale · A. Bharathi · M. Williams · N. M. Boopathi · M. Raveendran (✉)
Department of Plant Biotechnology, Centre for Plant Molecular Biology & Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India
e-mail: biotech@tnau.ac.in

N. M. Boopathi et al. (eds.), *The Moringa Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-80956-0_11

transcriptome sequencing of Moringa was decoded very recently (Tian et al. 2015; Chang et al. 2019; Panes et al. 2015). Such sequence information on medicinal plants will assist the understanding of biosynthesis pathways of medicinally important compounds.

Advances in sequencing technologies are evolving very rapidly from Sanger sequencing to long-read sequencing technologies. Next generation sequencing (NGS) platforms are mainly distinguished by their read length for instance. Second-generation sequencing technologies are capable of generating shorter read lengths (35–600 bp) compared to third-generation or long-read sequencing technologies (>1 kb). The major demerits of second-generation sequencing include data assembly and the inability to handle repeat sequences or large genomic rearrangements (Rhoads and Au 2015). The limitations of second-generation sequencing technologies can be overcome by long-read technologies as they can generate reads up to 2 Mb. Third-generation sequencing technologies including Pacbio, Nanopore, synthetic long reads, optical mapping, RNA-seq, hybrid sequencing (both second and third-generation techniques), and single-cell RNA/DNA sequencing are being widely adopted in the current decade. However, more advances in sequencing technologies are being evolved including advances in nanopore sequencing, in situ nucleic acid sequencing, microscopy-based sequencing (Kumar et al. 2019). The application of these advanced sequencing technologies has assisted the improvement of genome assemblies and transcriptome studies in many crops and also they served as a platform for analyzing underutilized crops. Such studies will promote the knowledge and understanding of the biochemical pathways of the important secondary metabolites and other chemical components in the medicinal crops resulting in dissecting the drug discovery mechanisms.

11.1.2 Third Generation Sequencing Technologies for Moringa Functional Genomics

Third-generation sequencing or long-read sequencing technologies hold numerous advantages compared to second-generation sequencing technologies. These long read technologies produce longer reads up to 10 kb compared to second-generation sequencing technologies which provide around ~600 bp read length (Amarasinghe et al. 2020). Moreover, third-generation sequencing technologies improve the efficiency of de novo assembly as second-generation sequencing technologies (such as HiSeq, MiSeq, NovaSeq, BGISEQ, and Ion torrent) involves challenges in genome assembly constitution with shorter reads (Dumschott et al. 2020).

Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies' (ONT) nanopore sequencing are the two major long-read sequencing technologies widely being adopted for crop species sequencing in the current sequencing era. SMRT involves sequencing by synthesis whereas the ONT approach involves a novel technique in which the individual DNA molecule moves through the pore and sensors detect the changes in the ionic current according to the passing nucleotide, and this information will be sued for base calling (Deamer et al. 2016). So far, ONT has been applied in a wide range of crop plants from model crops to non-model crops, and it benefits genomes with highly repetitive regions. Notable species sequenced using the ONT platform are Arabidopsis (Michael et al. 2018), rice (Mondal et al. 2018; Read et al. 2020; Tanaka et al. 2020; Choi et al. 2020), teak (Yasodha et al. 2018), sorghum (Deschamps et al. 2018), yam (Siadjeu et al. 2020), brassica (Belser et al. 2018), and tomato (Schmidt et al. 2017).

The first reference genome for duckweed (*S. intermedia*) was developed using Pacbio, and ONT platform and genomic sequences were compared with its sister species *S. polyrhiza*. Both species revealed more than 20,000 putative protein-coding genes, very low rDNA copy numbers, and a low amount of repetitive sequences, mainly Ty3/gypsy retroelements. This study also detected few new small chromosome rearrangements between both Spirodela species which refined the karyotype and the chromosomal sequence assignment for *S. intermedia* (Hoang et al. 2020). One of the major limitations of these long-read technologies is the high error rate compared to second-generation sequencing technologies, and it requires high-quality DNA.

However, considering the advantages over previous generation sequencing technologies third-generation sequencing technologies will become an essential sequencing tool with its upcoming rapid advances in plant genomics especially for large and complex genomes (Dumschott et al. 2020). As outlined in Fig. 11.1, a large array of functional genomics tools is available to harness maximum biological information from the plants'cells and these tools are continuously evolving.

11.1.3 Pan Genome Studies for Functional Genomics

In the genome sequencing era, the concept of a single reference genome will not be efficient as the single genotype may not represent the complete genome information of a species. Pan-genome is a core genome that represents the total gene count of a species including copy number variations (CNVs), presence/absence variations (PAVs), and SNPs (Munir et al. 2020). Initial methods involved assembling the whole genomes of various individuals and comparing the sequence variation, whereas the current methods involve identification of PAVs and non-assembling reads of individuals are assembled and added to pan-genome; recent methods also includes graph-based assembly which describes conservation and diversity in a species

The ever-increasing sequencing projects for the important crop species demands the need for pan-genome studies. The first pan-genome sequence was demonstrated in bacteria (Tettelin et al. 2005). In recent times, pan-genome studies play a major role in functional genomics. The first plant pan-genome assembly was made for soybean (Li et al. 2014). This study involved

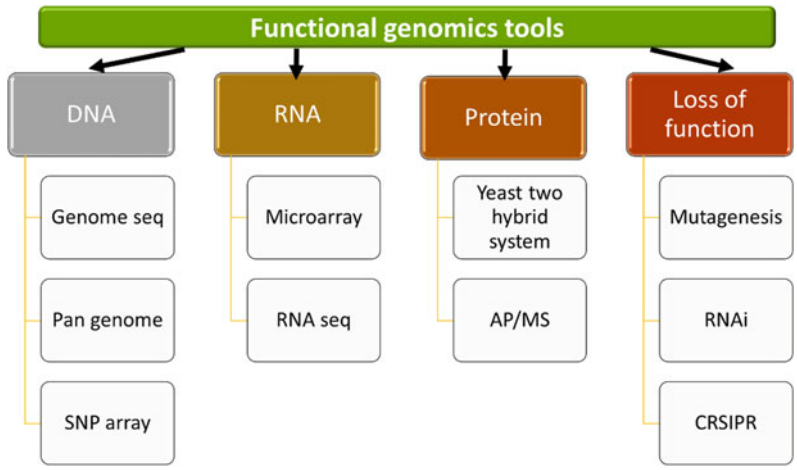

Fig. 11.1 Various functional analysis tools. (1) DNA level analysis which includes genome sequencing and structural analysis (2) RNA level analysis which includes transcriptome analysis and gene expression studies (3) Protein level analysis which includes proteomic studies revealing an entire set of protein in a species and their role (4) Mutation studies which include the creation of artificial mutation to study the function of a gene including Mutagenesis and CRISPR gene editing for functional analysis

core genome analysis of seven wild individuals resulting in the identification of genes with copy number variation, large-effect mutations, and positive selection relating to variations in agronomic traits including biotic stress, seed composition, flowering and maturity, organ size, and biomass.

Pan-genomes are being made available for many important crop species including Arabidopsis (Gan et al. 2011), rice (Schatz et al. 2014; Yao et al. 2015; Zhao et al. 2018; Wang et al. 2018), Brassica (Lin et al. 2014), soybean (Li et al. 2014; Liu et al. 2020), maize (Hirsch et al. 2014), wheat (Montenegro et al. 2017), capsicum (Ou et al. 2018), tomato (Gao et al. 2019), and walnut (Trouern-Trend et al. 2020).

Zhao et al. (2018) developed a pan-genome of 66 rice accessions belonging to *Oryza rufipogon* and *Oryza sativa*. This study identified 23 million sequence variations in the rice genome. They also reported the functional variants of five important quantitative trait loci (QTLs)—Hd3a (Os06g0157700), COLD1 (Os04g0600800), GW6a (Os06g0650300), TAC1 (Os09g0529300), and Sd1 (Os01g0883800), which are involved in flowering time, cold tolerance, grain weight, tiller angle, and plant height, respectively. A recent study in soybean reported the assembly of graph-based single reference genome constructed using 26 soybean accessions (Liu et al. 2020). This study identified large structural variations and gene fusion events and their functional role in gene expressions and agronomic traits. Pan-genome of tomato comprising 725 accessions revealed the absence of 4,873 genes from the reference genome (Gao et al. 2019). This study also revealed the role of *TomLoxC* promotor in controlling fruit flavor. Such pan-genome studies in important crop plants are becoming the source of variations and provide insights into domestication which will greatly assist breeding for better agronomic traits. However, pan-genome studies in Moringa have not yet been reported.

11.1.4 Long Read Sequencing Technologies for Transcriptomic Studies

Illumina's short-read sequencing platform, RNA-Seq has been popularly used for plant transcriptomic studies in the last decade. As short-read sequencing involves fragmentation of the full-length cDNA, loss of information from the original transcripts becomes unavoidable, and it is one of the major drawbacks of short-read sequencing platforms (Cui et al. 2020). Long read-sequencing platforms such as PacBio, ONT can sequence the full-length cDNA thus overcoming the loss of information which helps in analyzing the post-transcriptional events. The long-read sequencing platforms have been widely adopted in animals and being applied in plant transcriptomic studies especially since many uncharacterized species are being explored. In order to apply both sequencing platforms and to compare the results of these sequencing technologies, many studies have applied both Illumina and PacBio and /or ONT sequencing platforms for transcriptomic studies.

Several economically valued plants such as ornamental crops are being sequenced, and the understanding of the important characteristics such as flower color, fragrance are being carried out in the recent past using long-read sequencing technologies. For instance, Huang et al. (2020) utilized Illumina and PacBio platforms to study the molecular mechanism leading to the variation in flower color in ornamental crabapple. This study identified 603 differentially expressed genes (DEGs), including 449 upregulated and 154 downregulated genes. The role of transcription factors related to anthocyanin synthesis was reported, and five genes related to anthocyanin transport and degradation were found to be highly expressed in red petals. *R. lapponicum* is an important ornamental crop that possesses high economic value worldwide. Recently, Jia et al.

(2020) reported a full-length transcriptome of *R. lapponicum*. Based on KEGG analysis, this study identified 96 transcripts coding for the enzymes associated with anthocyanin synthesis.

Similarly, Zhang et al. (2019a) employed both Illumina and PacBio platforms in order to study the transcripts of brown planthopper (BPH) which is one of the major pests of rice. This study revealed the full diversity and complexity of the BPH transcriptome and indicates that BPH responses to rice resistance might be related to starvation stress responses, nutrient transformation, oxidative decomposition, and detoxification. PacBio sequencing has also been applied for transcriptomic studies of other important food crops including wheat (Dong et al. 2015), maize (Zhou et al. 2018; Wang et al. 2016), sorghum (Abdel-Ghany et al. 2016), coffee (Cheng et al. 2017), and garlic (Chen et al. 2018).

11.2 Advances in Moringa Functional Genomics

The high nutritional content of *M.oleifera* such as high protein, vitamin, and mineral content makes it attractive and widespread in developing countries. Every part of *M.oleifera* possess medicinal values, and it has been extensively studied for its nutritional and medicinal importance in treating diseases. For instance, anti-inflammatory, antimicrobial properties of root, leaf, stem, flower, and pod are reported to lower back pain, antidiabetic, and anti-cancer (Goyal et al. 2007). It is highly adapted to arid and semi-arid tropics due to its drought tolerance.

Genome sequencing and transcriptome sequencing of medicinal crops are being made available, and the pharmacological impacts are analyzed in detail for drug discovery. With the advent of NGS technologies, moringa draft genome and transcriptome are made available to explore the possibilities of understanding and enhancing its stress tolerance, nutritional, and pharmaceutical mechanisms.

The first transcriptome study on Moringa pods was demonstrated by Panes et al. (2015). This study generated a total of 182,588 transcripts out of which 3,556 unigenes were found to be involved in oil biosynthesis. Moreover, most of the unigenes were found to be involved in fatty acid biosynthesis with 1,009 unigenes, fatty acid catabolism with 982 unigenes, and triacylglycerol catabolism with 608 unigenes. Around 33 unigenes encodings for transcription factors were reported to be involved in regulating oil biosynthesis gene expression. This transcriptome resource for the *M. oleifera* Lam. mature seed embryo would assist in mapping of oil biosynthesis-related genes and the understanding of metabolic pathways which could possibly be used to improve seed yield and oil content of *M. oleifera*.

Draft genome of *M. oleifera* was reported by Chang et al. (2019) along with four other agriculturally important plants. This study predicted 18,451 protein-coding genes. Gene expansion and contraction study assisted the characterization of root nodule symbiosis genes, transcription factors, and starch biosynthesis-related genes in the five genomes. Moringa seeds are can be a good source of edible oil as the seeds are capable of producing oil up to 40%.

Pasha et al. (2020) reported the transcriptome of leaf, root, stem, seed, and flower of moringa cultivar, Bhagya. This study predicted 17,148 gene models and candidate genes related to the biosynthesis of secondary metabolites, vitamins, and ion transporters were identified. They also performed expression analysis through RT-PCR and metabolite quantification which showed a high expression pattern in the leaves, flowers, and seeds of the genes/enzymes involved in the biosynthesis of vitamins and metabolites like quercetin and kaempferol. In addition, this study revealed the expression of iron transporters and calcium storage proteins were observed in root and leaves and concluded that leaves retain the highest number of small molecules of interest. In continuation, assessment of the combined transcriptome for transcript abundance across five tissues assisted the prediction of the protein-coding genes. Further, these identified protein-coding genes from the transcripts were annotated and used for orthology analysis (Shafi et al. 2020).

Gene family evolution in Moringa was reported very recently by Lopez et al. (2020). This study reported gene expansion of 101 gene families grouping 957 genes, and the expanded families were highly enriched for chloroplastic and photosynthetic functions. In addition, this study also reported the large regions of plastid DNA (4.71%) insertions into the nuclear genome through microsynteny analysis of ten other plant species including rice, maize, and Arabidopsis.

11.2.1 Characteristics of Moringa Plastid Genome

The complete chloroplast genome of *M. oleifera* was reported by Liu et al. (2019a). The reported chloroplast genome size is 160,599 bp long and includes 113 full-length genes including 79 protein-coding genes. Its large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions are 88,933, 19,482, and 26,092 bp long, respectively. Phylogenetic tree analysis exhibited that *M. oleifera* was clustered with other Moringaceae species with 100% bootstrap values.

Another study by Mu et al. (2019) reported plastid genome sequence for the four species belonging to brassicales family including *M. oleifera*. In comparison to Liu et al. (2019a), the size of the plastid genome of *M. oleifera* was 163,131 bp and possess typical quadripartite structure: IRs, LSC, and SSC. The length of LSC, IR, and SSC regions of *M. oleifera* Lam. are 102,342 bp; 3,710 bp; 48,715 bp, respectively. All four species had the same number of 78 protein-coding genes, four ribosomal RNAs, while the number of transfer RNAs varies from 36 (*Cleome rutidosperma* DC.), 37 (*Carica papaya* L. and *Moringa oleifera* Lam.) to 38 (*Capparis urophylla* F.Chun).

More recently, Lin et al. (2019b) reported the complete chloroplast genome which was very close to the previously reported chloroplast genomes by Liu et al. (2019a) and Mu et al. (2019). The chloroplast genome was 160,600 bp in length with 88,577 bp of LSC, 18,883 bp of SSC, and 26,570 bp of IR. This study predicted 131 genes, and the phylogenetic analysis of 71 protein-coding sequences of 13 plant plastomes showed that the *M. oleifera* is closest to *Carica papaya*.

11.2.2 Functional Studies in Moringa

Availability of genome sequencing data provides opportunities to explore functional studies such as synteny, evolutionary and phylogenetic analysis using bioinformatics tools. Deng et al. (2016) screened 18 candidate genes selected from the Moringa transcriptome database. Expression stabilities of the selected agronomically important traits were examined in 90 samples collected from the pods in different developmental stages, various tissues, and the roots and leaves under different conditions (low or high temperature, sodium chloride (NaCl)- or polyethyleneglycol (PEG)- simulated water stress).

This study provided insights on the Moringa genes involved in abiotic stress tolerance and forms a basis for Moringa functional gene analysis. The first proteomic analysis of the flower of *M. oleifera* was reported by Shi et al. (2018). This study identified 9443 peptides corresponding to 4004 high-confidence proteins and a number of commercially important food-grade enzymes were also commented. A total of 261 proteins were annotated as carbohydrate-active enzymes, 16 proteases, 22 proteins are assigned to the citrate cycle, which the top proteins were assigned to the GH family, cysteine synthase, and serine/threonine-protein phosphatase. These enzymes indicated that they are a new source with potential use for fermentation and brewing industry, fruit and vegetable storage, and the development of functional peptides.

WRKY transcription factors are known to be involved in numerous plant processes from germination to senescence. With the resource available on the Moringa genome (Tian et al. 2015), genome-wide identification and characterization of WRKY transcription factors were reported by Zhang et al. (2019b). This study identified 54 MoWRKY TFs, and the expression

analysis using RT-PCR revealed the involvement of potential MoWRKY genes with respect to abiotic stresses such as salt, heat, drought, H_2O_2, and cold.

A recent study reported the genetic diversity of 57 M. *oleifera* accessions using RAPD marker and their biologically active component such as cinnamic, caffeic, ferulic, and coumaric acids, flavonoids analysis using HPLC (Panwar and Mathur 2020). This study grouped the 57 accessions into five groups, and high diversity in the concentration of active compounds was also reported using HPLC. The strong correlation between phytochemical variables and DNA polymorphism will assist in breeding for selecting the best accessions.

11.2.3 Metabolomics in Moringa

As moringa possess versatile utility for medicine and nutrition sources, profiling of its metabolites and making the common metabolite database of various tissues are very important for functional studies. Mahmud et al. (2014) reported the profiling of metabolites from moringa leaf and stem tissues. Among the 30 metabolites identified in this study, 22 metabolites were common in both leaf and stem tissues and the remaining eight metabolites included, 4-aminobutyrate, adenosine, guanosine, tyrosine, and p-cresol were found only in leaf tissues; whereas, glutamate, glutamine, and tryptophan were found only in stem tissues. They also performed biochemical pathway analysis which revealed that 28 identified metabolites were interconnected with 36 different pathways as well as related to different fatty acids and secondary metabolites synthesis biochemical pathways.

Flavonoids are important secondary metabolites with specific metabolic functions in plants. *Moringa oleifera* and *M. ovalifolia* are two moringa species known to contain a wide spectrum of flavonoids molecules with known nutraceutical properties. A comparative analysis of flavonoid content in *M. oleifera* and *M. ovalifolia* with the aid of ultra-high-performance liquid chromatography coupled with high-resolution quadrupole time-of-flight mass spectrometer (UHPLC-qTOF-MS) fingerprinting was demonstrated by Makita et al. (2016). Various flavonoids identified from these two species conclude that the various genetic bases of flavonoid biosynthesis in these species. The differentiation of the flavonoids among these species was mainly due to the superior glycosylation capabilities of *M. olifera* compared *M. ovalifolia*. This study concluded that *M. olifera* has wide pharmacological application based on its glycosylation complexity.

The metabolite and protein content of the plant are highly influenced by soil types. A recent study performed metabolite profiling of moringa leaves cultivated with vermicompost and phosphate rock under water stress conditions (Albores et al. 2019). UPLC-ESI–MS/MS analysis of leaf extracts revealed that the most abundant metabolites were flavonoids, alkaloids, and terpenes. This study identified that the water stress-induced changes in the metabolomics profile and the morphometric variables of *M. oleifera*.

A comparative study of the chemical constituents from moringa leaves collected from different cultivation regions, i.e., India and China were reported using liquid chromatography and mass spectroscopy (Lin et al. 2019c). This study reported a total of 122 characterized components, containing 118 shared constituents, from moringa leaves of India and China. Such a comprehensive phytochemical profile study provides insights into the basis for explaining the effect of different growth environments on secondary metabolites.

A very recent study by Rocchetti et al. (2020) had comprehensively investigated the (poly)-phenolic profile of *M. oleifera* leaves through untargeted metabolomics, following a homogenizer-assisted extraction (HAE) using three solvent systems, i.e., methanol (HAE-1), methanol–water 50:50 v/v (HAE-2), and ethyl acetate (HAE-3). They annotated close to 300 compounds, recording mainly flavonoids and phenolic acids. In addition, they also reported antioxidant capacity, antimicrobial activity, and enzyme inhibition assays in the different extracts.

This study concluded that *M. oleifera* leaf extracts are a good source of bioactive polyphenols with potential use in the food and pharma industries.

11.3 Medicinal and Pharmaceutical Studies on *M. oleifera*

Several studies have reported the benefits of the Moringa plant (leaf, flower, seed, stem) and its extracts in controlling non-communicable diseases such as diabetes, obesity, cancer, heart disease, and stroke (Lin et al. 2018). Though a smaller number of studies are available on humans, many studies on animals have reported a positive association between the consumption of *M. oleifera*-containing foods and a reduced risk of developing certain types of NCDs. For instance, Li et al. (2020) reported transcriptome gene expression and epigenome DNA methylation in mouse kidney mesangial cells (MES13) using next-generation sequencing technology. After high glucose treatment, epigenome and transcriptome were found to be significantly altered and exposure to Moringa isothiocyanate (MIC-1) which is a bioactive constituent found abundantly in *M. oleifera*, possesses antioxidant and anti-inflammation properties reversed some of the changes caused by high glucose.

Another study by Cheng et al. (2019) reported the *Nrf2-ARE* antioxidant activity of MIC-1, and its potential in diabetic nephropathy. In brief, this study concluded that MIC-1 activates *Nrf2-ARE* signaling, increases expression of *Nrf2* target genes, and suppresses inflammation, while also reducing oxidative stress and possibly *TGFβ1* signaling in high glucose-induced renal cells. Sun et al. (2019) showed the positive effects of Moringa leaf extract on the treatment of type 2 diabetes mellitus by influencing the expression *Per1* and *Per2* genes. This study concluded that moringa leaf extract can ameliorate liver damage in diabetic rats, possibly due to its anti-glycation activities. In addition, anti-inflammatory, antioxidative, and anti-cancer properties of MIC-1 were reported by Wang et al. (2019).

Natural plant-derived biostimulants are proven to improve the growth, yield, and post-harvest quality of horticultural products. Moringa leaf extracts in particular have been shown to improve seed germination, plant growth and yield, nutrient use efficiency, crop, and product quality traits (pre- and post-harvest), as well as tolerance to abiotic stresses (Zulfiqar et al. 2020). The use of plant-derived biostimulants such as moringa leaf extracts can help in reducing the fertilizer quantities needed and thus contribute to achieving global food security sustainably.

11.4 Future perspectives

Though the medicinally important moringa crop is being explored in the recent past, there is a need to improve its genomics resources such as markers, QTLs, and candidate gene identification for crop improvement. It is important to explore genome-wide analysis with an increased number of accessions for diversity studies and understanding potential marker-trait association. Such studies with available third-generation sequencing technologies will provide millions of SNPs leading to the development of SNP-chip arrays which can serve as rich marker resources especially for marker-assisted selection. Genome sequence of various accessions of moringa will pave way for developing pan-genomes which will assist the identification of sequence-level variation such as CNVs and PAVs. CNVs have potential effects on gene expression and structure relating to phenotypic changes in various accessions. Developing the core genome of moringa will identify the potential candidate genes involving in its abiotic stress-tolerant mechanism and biochemical pathways of important secondary metabolites contributing to its anti-inflammatory properties.

References

Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS (2016) A survey of the sorghum transcriptome using single-molecule long reads. Nat Commun 7(1):1–11

Al-Asmari AK, Albalawi SM, Athar MT, Khan AQ, Al-Shahrani H, Islam M (2015) Moringa oleifera as an anti-cancer agent against breast and colorectal cancer cell lines. PloS One 10(8):e0135814

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21(1):1–16

Anwar F, Latif S, Ashraf M, Gilani AH (2007) Moringa oleifera: a food plant with multiple medicinal uses. Phytother Res Int J Devoted Pharmacol Toxicol Eval Nat Prod Deriv 21(1):17–25

Belser C, Istace B, Denis E, Dubarry M, Baurens FC, Falentin C, Genete M, Berrabah W, Chèvre AM, Delourme R, Deniot G (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants 4(11):879–887

Chang Y, Liu H, Liu M, Liao X, Sahu SK, Fu Y, Song B, Cheng S, Kariba R, Muthemba S, Hendre PS (2019) The draft genomes of five agriculturally important African orphan crops. GigaScience 8(3):giy152

Chen X, Liu X, Zhu S, Tang S, Mei S, Chen J, Li S, Liu M, Gu Y, Dai Q, Liu T (2018) Transcriptome-referenced association study of clove shape traits in garlic DNA Res 25(6):587–596

Cheng B, Furtado A, Henry RJ (2017) Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. Gigascience, 6 (11):gix086

Cheng D, Gao L, Su S, Sargsyan D, Wu R, Raskin I, Kong AN (2019) Moringa isothiocyanate activates Nrf2: potential role in diabetic nephropathy. AAPS J 21(2):31

Choi JY, Lye ZN, Groen SC, Dai X, Rughani P, Zaaijer S, Harrington ED, Juul S, Purugganan MD (2020) Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. Genome Biol 21(1):21

Cui J, Lu Z, Xu G, Wang Y, Jin B (2020) Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the Arabidopsis transcriptome. Plant Methods 16(1):1–13

Deamer D, Akeson M, Branton D (2016) Three decades of nanopore sequencing. Nat Biotechnol 34(5):518–524

Deng LT, Wu YL, Li JC, OuYang KX, Ding MM, Zhang JJ, Li SQ, Lin MF, Chen HB, Hu XS, Chen XY (2016). Screening reliable reference genes for RT-qPCR analysis of gene expression in Moringa oleifera. PloS One 11(8):e0159458

Deschamps S, Zhang Y, Llaca V, Ye L, Sanyal A, King M, May G, Lin H (2018) A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. Nat Commun 9 (1):1–10

Dong L, Liu H, Zhang J, Yang S, Kong G, Chu JS, Chen N, Wang D (2015) Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. BMC Genomics 16(1):1039

Dumschott K, Schmidt MHW, Chawla HS, Snowdon R, Usadel B (2020) Oxford nanopore sequencing: new opportunities for plant genomics? J Exp Bot

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, Kahles A (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477(7365):419–423

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL, Thannhauser TW (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51(6):1044–1051

Goyal BR, Agrawal BB, Goyal RK, Mehta AA (2007) Phyto-pharmacology of Moringa oleifera Lam.—an overview. Nat Prod Radiance 6:347–353

Guzmán-Albores JM, Ramírez-Merchant ML, Interiano-Santos EC, Manzano-Gómez LA, Castañón-González JH, Winkler R, Abud-Archila M, Montes-Molina JA, Gutiérrez-Miceli FA, Ruíz-Valdiviezo VM (2019) Metabolomic and proteomic analysis of Moringa oleifera cultivated with vermicompost and phosphate rock under water stress conditions. Int J Agric Biol 21 (4):786–794

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26(1):121–135

Hoang PT, Fiebig A, Novák P, Macas J, Cao HX, Stepanenko A, Chen G, Borisjuk N, Scholz U, Schubert I (2020) Chromosome-scale genome assembly for the duckweed Spirodela intermedia, integrating cytogenetic maps, PacBio and Oxford Nanopore libraries. Sci Rep 10(1):1–14

Huang B, Rong H, Ye Y, Ni Z, Xu M, Zhang W, Xu LA (2020) Transcriptomic analysis of flower color variation in the ornamental crabapple (Malus spp.) half-sib family through Illumina and PacBio Sequel sequencing. Plant Physiol Biochem 149:27–35

Jamnadass R, Mumm RH, Hale I, Hendre P, Muchugi A, Dawson IK, Powell W, Graudal L, Yana-Shapiro H, Simons AJ, Van Deynze A (2020) Enhancing African orphan crops with genomics. Nat Genet 52(4):356–360

Jia X, Tang L, Mei X, Liu H, Luo H, Deng Y, Su J (2020) Single-molecule long-read sequencing of the full-length transcriptome of Rhododendron lapponicum L. Sci Rep 10(1):1–11

Kumar KR, Cowley MJ, Davis RL (2019) Next-generation sequencing and emerging technologies. In: Seminars in thrombosis and hemostasis, vol 45, no 07. Thieme Medical Publishers, pp 661–673. (2019, October)

Li S, Li W, Wu R, Yin R, Sargsyan D, Raskin I, Kong AN (2020) Epigenome and transcriptome study of moringa isothiocyanate in mouse kidney mesangial cells induced by high glucose, a potential model for diabetic-induced nephropathy. AAPS J 22(1):8

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32(10):1045–1052

Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RG, Wang X, de Ridder D, Bonnema G (2014) Beyond genomic variation-comparison and functional annotation of three Brassica rapagenomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics 15(1):250

Lin M, Zhang J, Chen X (2018) Bioactive flavonoids in Moringa oleifera and their health-promoting properties. J Funct Foods 47:469–479

Liu J, Cai HH, Li HQ, Liu ZY, Zheng C, Shi C, Niu YF (2019a) The chloroplast genome of Moringa oleifera (Moringaceae). Mitochondrial DNA Part B 4(1):646–647

Lin W, Dai S, Chen Y, Zhou Y, Liu X (2019b). The complete chloroplast genome sequence of Moringa oleifera Lam.(Moringaceae). Mitochondrial DNA Part B, 4(2):4094–4095

Lin H, Zhu H, Tan J, Wang H, Wang Z, Li P, Zhao C, Liu J (2019c) Comparative analysis of chemical constituents of Moringa oleifera leaves from China and India by ultra-performance liquid chromatography coupled with quadrupole-time-of-flight mass spectrometry. Molecules 24(5):942

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, Huang X (2020) Pan-genome of wild and cultivated soybeans. Cell 182(1):162–176

Mahmud I, Chowdhury K, Boroujerdi A (2014) Tissue-specific metabolic profile study of Moringa oleifera L. using nuclear magnetic resonance spectroscopy.In: Plant tissue culture & biotechnology. Bangladesh Association for Plant Tissue Culture & Biotechnology; BAPTC&B, vol 24, no 1, pp 77

Makita C, Chimuka L, Steenkamp P, Cukrowska E, Madala E (2016) Comparative analyses of flavonoid content in Moringa oleifera and Moringa ovalifolia with the aid of UHPLC-qTOF-MS fingerprinting. S Afr J Bot 105:116–122

Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat Commun 9 (1):1–8

Mondal TK, Rawal HC, Chowrasia S, Varshney D, Panda AK, Mazumdar A, Kaur H, Gaikwad K, Sharma TR, Singh NK (2018) Draft genome sequence of first monocot-halophytic species Oryza coarctata reveals stress-specific genes. Sci Rep 8(1):1–13

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. Plant J 90(5):1007–1013

Mu W, Yang T, Liu X (2019) The complete plastid genomes of four species from Brassicales. Mitochondrial DNA Part B 4(1):124–125

Munir F, Saba NU, Arveen M, Siddiqa A, Ahmad J, Amir R (2020) Pan-genomics of plants and its applications. In: Pan-genomics: applications, challenges, and future prospects. Academic Press, pp 285–306

Ojeda-López J, Marczuk-Rojas JP, Polushkina OA, Purucker D, Salinas M, Carretero-Paulet L (2020) Evolutionary analysis of the Moringa oleifera genome reveals a recent burst of plastid to nucleus gene duplications. Sci Rep 10(1):1–15

Ou L, Li D, Lv J, Chen W, Zhang Z, Li X, Yang B, Zhou S, Yang S, Li W, Gao H (2018) Pan-genome of cultivated pepper (Capsicum) and its use in gene presence–absence variation analyses. New Phytol 220 (2):360–363

Panes VA, Kitazumi A, Butler M, Baoas R, De los Reyes BG (2015) Analysis of the oil biosynthesis transcripts of the Moringa oleifera Lam. mature seed embryos using RNA sequencing. In:I international symposium on Moringa, vol 1158, pp 55–62 (2015, November)

Panwar A, Mathur J (2020) Genetic and biochemical variability among Moringa oleifera Lam. accessions collected from different agro-ecological zones. Genome 63(3):169–177

Pasha SN, Shafi KM, Joshi AG, Meenakshi I, Harini K, Mahita J, Sajeevan RS, Karpe SD, Ghosh P, Nitish S, Gandhimathi A (2020) The transcriptome enables the identification of candidate genes behind medicinal value of Drumstick tree (Moringa oleifera). Genomics 112(1):621–628

Ramachandran C, Peter KV, Gopalakrishnan PK (1980) Drumstick (Moringa oleifera): a multipurpose Indian vegetable. Econ Bot 276–283

Read AC, Moscou MJ, Zimin AV, Pertea G, Meyer RS, Purugganan MD, Leach JE, Triplett LR, Salzberg SL, Bogdanove AJ (2020) Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. *PLoS genetics*, *16*(1), e1008571.

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics Proteomics Bioinform 13 (5):278–289

Rocchetti G, Pagnossa JP, Blasi F, Cossignani L, Piccoli RH, Zengin G, Montesano D, Cocconcelli PS, Lucini L (2020) Phenolic profiling and in vitro bioactivity of Moringa oleifera leaves as affected by different extraction solvents. Food Res Int 127:108712

Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, Wright MH (2014) Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. Genome Biol 15(11):506

Schmidt MH, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, Bolger ME, Alseekh S, Maß J, Pfaff C, Schurr U (2017) De novo assembly of a new Solanum pennellii accession using nanopore sequencing. Plant Cell 29(10):2336–2348

Sehgal N, Gupta A, Valli RK, Joshi SD, Mills JT, Hamel E, Khanna P, Jain SC, Thakur SS, Ravindranath V (2012) Withania somnifera reverses Alzheimer's disease pathology by enhancing low-density lipoprotein receptor-related protein in liver. Proc Natl Acad Sci 109(9):3510–3515

Shafi KM, Joshi AG, Meenakshi I, Pasha SN, Harini K, Mahita J, Sajeevan RS, Karpe SD, Ghosh P, Nitish S, Gandhimathi A (2020) Dataset for the combined transcriptome assembly of M. oleifera and functional annotation. Data in Brief, 105416.

Shi Y, Wang X, Huang A (2018) Proteomic analysis and food-grade enzymes of Moringa oleifer Lam. a Lam. flower. Int J Biol Macromol 115:883–890

Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B (2020) High contiguity de novo genome sequence assembly of Trifoliate yam (Dioscorea dumetorum) using long read sequencing. Genes 11 (3):274

Sun W, Liu J, Wu L, Guo X, Zhang L, Fan Y, Yang L, Guo X, Hou Y, Mu X, Qin L (2019) Transcriptome analysis of the effects of Moringa Oleifera leaf extract in db/db mice with type 2 diabetes mellitus. Int J Clin Exp Med 12(6):6643–6658

Tanaka T, Nishijima R, Teramoto S, Kitomi Y, Hayashi T, Uga Y, Kawakatsu T (2020) De novo genome assembly of the indica rice variety IR64 using linked-read sequencing and nanopore sequencing. G3: Genes, Genomes, Genetics 10(5):1495–1501

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." Proc Natl Acad Sci 102(39):13950–13955

Tian Y, Zeng Y, Zhang J, Yang C, Yan L, Wang X, Shi C, Xie J, Dai T, Peng L, Huan YZ (2015) High quality reference genome of drumstick tree (Moringa oleifera Lam.), a potential perennial crop. Sci China Life Sci 58(7):627–638

Trouern-Trend AJ, Falk T, Zaman S, Caballero M, Neale DB, Langley CH, Dandekar AM, Stevens KA, Wegrzyn JL (2020) Comparative genomics of six Juglans species reveals disease-associated gene family contractions. Plant J 102(2):410–423

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat Commun 7(1):1–13

Wang C, Wu R, Sargsyan D, Zheng M, Li S, Yin R, Su S, Raskin I, Kong AN (2019) CpG methyl-seq and RNA-seq epigenomic and transcriptomic studies on the preventive effects of Moringa isothiocyanate in mouse epidermal JB6 cells induced by the tumor promoter TPA. J Nutr Biochem 68:69–78

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557(7703):43–49

Yao W, Li G, Zhao H, Wang G, Lian X, Xie W (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biol 16 (1):1–20

Yasodha R, Vasudeva R, Balakrishnan S, Sakthi AR, Abel N, Binai N, Rajashekar B, Bachpai VK, Pillai C, Dev SA (2018) Draft genome of a high value tropical timber tree, Teak (Tectona grandis L. f): insights into SSR diversity, phylogeny and conservation. DNA Res 25(4):409–419

Zhang J, Guan W, Huang C, Hu Y, Chen Y, Guo J, Zhou C, Chen R, Du B, Zhu L, Huanhan D (2019a) Combining next-generation sequencing and single-molecule sequencing to explore brown plant hopper responses to contrasting genotypes of japonica rice. BMC Genomics 20(1):1–18

Zhang J, Yang E, He Q, Lin M, Zhou W, Pian R, Chen X (2019b). Genome-wide analysis of the WRKY gene family in drumstick (Moringa oleifera Lam.). Peer J 7: e7063

Zhang J, Yang E, He Q, Lin M, Zhou W, Pian R, Chen X (2019c) Genome-wide analysis of the WRKY gene family in drumstick (Moringa oleifera Lam.). Peer J 7: e7063

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, Wang Y (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet 50 (2):278–284

Zhou Y, Zhao Z, Zhang Z, Fu M, Wu Y, Wang W (2019) Isoform sequencing provides insight into natural genetic diversity in maize. Plant Biotechnol J 17 (8):1473

Zulfiqar F, Casadesús A, Brockman H, Munné-Bosch S (2020) An overview of plant-based natural biostimulants for sustainable horticulture with a particular focus on moringa leaf extracts. Plant Sci 295:110194