

Supplementary Table S1 Reported properties, applications and uses of *Moringa oleifera* plant parts.

<i>Moringa oleifera</i> Leaves	Properties	▪ Antioxidant ▪ Flavonoids ▪ Minerals and Vitamins ▪ Amino acids	(Gopalakrishnan <i>et al.</i> , 2016; Gopalakrishnan <i>et al.</i> , 2016; Oyeyinka and Oyeyinka, 2018; Vergara-Jimenez <i>et al.</i> , 2017)
	Applications	▪ Food supplements ▪ Source of proteins	
	Benefits	▪ Therapeutic ▪ Stabilizes cell structure ▪ Promotes metabolism	
<i>Moringa oleifera</i> seeds	Properties	▪ Antioxidant ▪ Anti-microbial ▪ Anti-inflammatory ▪ Phenolic ▪ Bio-active ▪ Compounds	(Brilhante <i>et al.</i> , 2017; Gopalakrishnan <i>et al.</i> , 2016)
	Applications	▪ Anti-fungal ▪ Oil extraction ▪ Water purification ▪ Lubricant ▪ Medicinal	
	Benefits	▪ Skin remedies ▪ Liver Health	
Ben oil (<i>M. oleifera</i>)	Properties	▪ Anti-aging ▪ Antioxidant ▪ Exfoliant ▪ Preservative ▪ Anti-inflammatory	(Vergara-Jimenez <i>et al.</i> , 2017; Pareek <i>et al.</i> , 2023)
	Applications	▪ High oxidation stability ▪ Soap ▪ Lubricant ▪ Food (Edible oil) ▪ Cosmetics (skin, hair, perfume base)	
	Benefits	▪ Skin remedies ▪ Rheumatic oil for arthritic joints ▪ Excellent moisturizer	

Supplementary Table S2 Sequencing summary for the *Moringa oleifera* (*M. oleifera*) genome (Illumina HiSeq, PE150). Raw and post-filtering (“clean”) statistics are shown. Read counts are in millions (M); bases are in gigabases (Gb); quality is reported as the percentage of bases with Q_≥20 (Q20) or Q_≥30 (Q30).

Metric	Raw reads	Clean reads
Read length (bp)	150 bp	149 bp
Read pairs (million)	173.11	172.12
Total bases (Gb)	124.20	121.40
Q20 bases (%)	97.1	97.3
Q30 bases (%)	92.1	92.5
GC content of reads (%)	37.77	37.59

Supplementary Table S3 Assembly statistics across ABySS k-mer values. QUAST metrics are shown for scaffolds $\geq 1,000$ bp. The selected build (k = 91) maximized contiguity (N50 = 17,279 bp) with the fewest scaffolds.

Assembly (k)	Longest scaffold	Scaffolds (≥ 1000 bp)	N50
31	974,48	23,545	6,340
71	312,914	28,711	11,614
91	395,443	13,872	17,279
121	318,204	35,488	14,614

Note: Statistics computed with QUAST v5.0.2 on scaffolds $\geq 1,000$ bp

Supplementary Table S4 Repeat content by class and subclass in the assembled genome of *Moringa oleifera*.

Repeat annotations were generated with RepeatMasker v4.1.0 using a species-specific library built by RepeatModeler v2.0.1 (RECON + RepeatScout). Statistics are computed on scaffolds ≥ 1 kb (total assembly length 205,248,313 bp; 13,872 sequences).

Class	Number of elements	Length occupied (bp)	Percentage of genome (%)
Retroelements	36483	32585386	15.88
SINEs	142	29487	0.01
LINEs	6971	6813393	3.32
LTR elements	29370	25742506	12.54
DNA transposons	23141	18617872	9.07
Unclassified- interspersed	88757	55245172	26.92
Total interspersed repeats	—	106448430	51.86
Rolling circles	410	321098	0.16
Small RNA	8772	10312333	5.02
Satellites	260	79406	0.04
Simple repeats	74014	2945867	1.44
Low complexity	18918	942157	0.46

Note: Percentages are relative to the assembly (205,248,313 bp). “Total interspersed repeats” = Retroelements + DNA transposons + Unclassified-interspersed. Other classes (e.g., simple repeats, low complexity) are listed separately and may overlap with interspersed repeats; therefore values need not sum to 100%.



Figure S1. Illumina read quality for the *Moringa oleifera* genome (PE150). (A–B) Per-base Phred scores for read 1 (A) and read 2 (B) show uniformly high quality across cycles (majority $\geq Q30$). (C–D) Per-base nucleotide composition profiles for read 1 (C) and read 2 (D) are stable after the initial cycles, consistent with random genomic libraries. (E–F) GC-content distributions for read 1 (E) and read 2 (F) are unimodal and centered near the assembly GC ($\approx 35\text{--}38\%$), indicating low contamination. Summary metrics: raw data ~ 124.2 Gb; clean data ~ 121.4 Gb; $\geq 92\%$ bases $Q \geq 30$.