Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

Plant Communications
Resource article

CellPress
Partner Journal

HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads

Shuyuan Tang[1,2], Yanyang Liang[2], Fengqi Wu[2], Ya Wu[3], Jiawei Li[1], Lvna Lin[2], Junting Feng[4], Aiping Luan[1], You Wang[1], Chunxiang Xu[2], Jiabao Wang[1,3], Yongfeng Zhou[1], Rui Xia[2], Yi Liao[2] and Chengjie Chen[1,*]

[1]State Key Laboratory of Tropical Crop Breeding, Key Laboratory of Crop Gene Resources and Germplasm Enhancement in South China, Ministry of Agriculture and Rural Affairs, Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China

[2]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Horticulture, South China Agricultural University, Guangzhou, Guangdong 510640, China

[3]Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China

[4]State Key Laboratory of Tropical Crop Breeding, Sanya Research Institute, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 572024, China

***Correspondence: Chengjie Chen (ccj@catas.cn)**

https://doi.org/10.1016/j.xplc.2025.101467

ABSTRACT

Chloroplast and mitochondrial genomes, which coexist with the nuclear genome, are extensively used in gene function studies, evolutionary analyses, and targeted breeding. However, assembling these organelle genomes remains challenging due to frequent recombination events and abundant repetitive sequences. Although some assembly tools are available, many are difficult to install or require extensive parameter tuning and computational resources. To address these limitations, we present a high-fidelity, data-driven mitochondrial genome assembly toolkit (HiMT), a user-friendly, out-of-the-box software solution that enables one-click chloroplast and mitochondrial genome assembly using default parameters, particularly for plant species. HiMT automatically estimates read coverage depth and employs a fixed *k*-mer prefix strategy to minimize computational demands, making it suitable for use on standard laptops. Benchmark tests show that HiMT delivers complete assemblies with a high success rate, fast runtimes, and low hardware requirements. Additionally, HiMT features a graphical user interface (GUI) and generates interactive reports for assessing the quality of organelle genome assemblies. We anticipate that HiMT will facilitate high-quality plant mitochondrial genome research and significantly streamline organelle genome assembly workflows. To support the research community, HiMT is freely available to non-commercial users at https://github.com/tang-shuyuan/HiMT.

Key words: mitochondria, plastids, genome assembly, algorithm optimization, software development

Tang S., Liang Y., Wu F., Wu Y., Li J., Lin L., Feng J., Luan A., Wang Y., Xu C., Wang J., Zhou Y., Xia R., Liao Y., and Chen C. (2025). HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads. Plant Comm. **6**, 101467.

INTRODUCTION

Eukaryotic organisms can possess up to three major genomes: the nuclear genome, the mitochondrial genome, and, for plants, the chloroplast genome, each containing essential genetic material and having specific biological functions (Wang et al., 2024). For all these genomes, the accuracy of reference sequences and gene loci is vital for gene function analyses, evolutionary research, and targeted breeding applications (Petit et al., 2004; Kang et al., 2021). However, the efficiency and ease of analyzing these genomes vary considerably.

Due to its critical role, larger size, and complexity, the nuclear genome remains the primary focus of most biological research (Sun et al., 2021). In contrast, the mitochondrial genome in animals and the chloroplast genome in plants are smaller and more conserved in sequence and structure. This conservation makes the assembly of these genomes relatively straightforward. As a result, animal mitochondrial and plant chloroplast genomes have been widely sequenced for diverse research purposes (Martin et al., 2005; Daniell et al., 2016). In comparison, plant mitochondrial genomes exhibit a high degree of complexity. Although they typically harbor conserved

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

Features	HiMT	PMAT	TIPPo	GetOrganelle	GSAT
Supporting data types	long reads	long reads	long reads	short reads	hybrid
CP assembly	✓	✓	✓	✓	×
Animal MT assembly	✓	×	×	✓	×
Plant MT assembly	✓	✓	✓	✓	✓
Cross-platform	✓	×	×	×	×
Conda package	✓	×	✓	✓	×
Container support	✓	×	✓	×	×
GUI	✓	×	×	×	×
Assembly graph visualization	✓	×	×	×	×
Quality assessment report	✓	×	×	×	×

Table 1. Feature comparison between HiMT and available organelle genome assembly tools

"Long reads" denotes HiFi reads or error-corrected Oxford Nanopore sequencing data. "MT" and "CP" denote mitochondrial and chloroplast genomes, respectively.

mitochondrial genes, their overall size and structure vary considerably due to evolutionary processes, leading to substantial differences even among closely related species (Wynn and Christensen, 2019; Wang et al., 2024). Therefore, the assembly and application of plant mitochondrial genomes remain relatively uncommon compared to nuclear and chloroplast genomes.

To assemble organelle genomes from whole-genome sequencing (WGS) reads, several software tools have been developed, including GetOrganelle, GSAT, PMAT, and, more recently, TIPPo (Jin et al., 2020; He et al., 2023; Bi et al., 2024; Xian et al., 2025). All these tools are based on two assembly strategies. The first strategy consists of assembling all available sequences using nuclear genome assembly tools such as hifiasm (Cheng et al., 2021) and NextDenovo (Hu et al., 2024), followed by the identification of contigs corresponding to plastids or mitochondrial genomes. This approach often yields a complete, though non-circularized, chloroplast or animal mitochondrial genome. However, for plant mitochondrial genomes, this often results in partial assemblies. Therefore, a "read, bait, and extend" process is commonly employed, consuming significant runtime and computational resources. The second strategy selects organelle genome reads and then applies circularized genome assembly tools, such as SPAdes (Bankevich et al., 2012) and Flye (Kolmogorov et al., 2019). The primary challenge of this approach lies in defining appropriate filtering rules that accurately identify organelle reads.

As mentioned above, the complexity and variability of plant mitochondrial genomes pose significant challenges for their assembly and analysis, especially when relying solely on short-read sequencing data (He et al., 2022; Bi et al., 2024; Xian et al., 2025). The recent development of PacBio HiFi sequencing has made it possible to generate highly accurate long reads, which can span complex recombination regions and thereby enhance mitochondrial genome assembly quality (Bi et al., 2024). Despite these advances, currently, only PMAT and TIPPo can integrate whole-genome PacBio HiFi sequencing data, using specific parameters (Bi et al., 2024; Xian et al., 2025). All these existing tools and workflows demand significant computational resources, involve complex environmental setups, and require parameter adjustments. They also require proficiency in command-line operations, limiting researchers' ability to efficiently obtain the reference sequences of plant mitochondrial genomes.

To address these challenges and limitations, we developed HiMT (High-Fidelity Data-based Mitochondrial Genome Assembly Toolkit), a software package that optimizes algorithm efficiency and minimizes computational resource consumption during organelle genome assembly. HiMT integrates all necessary dependencies and is built for cross-platform functionality, particularly for Windows systems, ensuring wide user accessibility. The tool also features a graphical user interface (GUI) developed using the TBtools-II platform (Chen et al., 2023). Additionally, HiMT provides interactive reports for assessing assembly quality and includes comparison tools for evaluating multiple assembly outputs, thereby filling a key gap in the current landscape of mitochondrial genome assembly quality evaluation tools (Table 1).

RESULTS

Optimization of algorithms and design of workflows

In plant WGS, organelle genomes are typically sequenced at higher depths than nuclear genomes, with mitochondrial genomes exhibiting over 5-fold and chloroplast genomes over 50-fold greater depth (Bi et al., 2024). Mitochondrial genomes contain 24 core protein-coding genes, a variable number of additional genes, tRNAs, and rRNAs (Mower et al., 2012). Based on these characteristics, we first assessed the sequencing depth of core mitochondrial genes using HiFi data. This depth was then used to estimate retrospectively the minimum *k*-mer coverage for mitochondrial and chloroplast genome-derived sequences, enabling the extraction of reads containing high-frequency *k*-mers.

Conventional assembly tools typically store all possible *k*-mers and calculate their depth, a computationally intensive and redundant process for HiFi reads, which are typically longer than 10 kb. In HiMT, we optimized this step by retaining only *k*-mers with a specific 3-bp prefix, allowing us to efficiently cover nearly all reads (Figure 1). This method significantly reduces computational complexity and memory usage by approximately 1/64 compared to global *k*-mer analysis and produces a streamlined

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

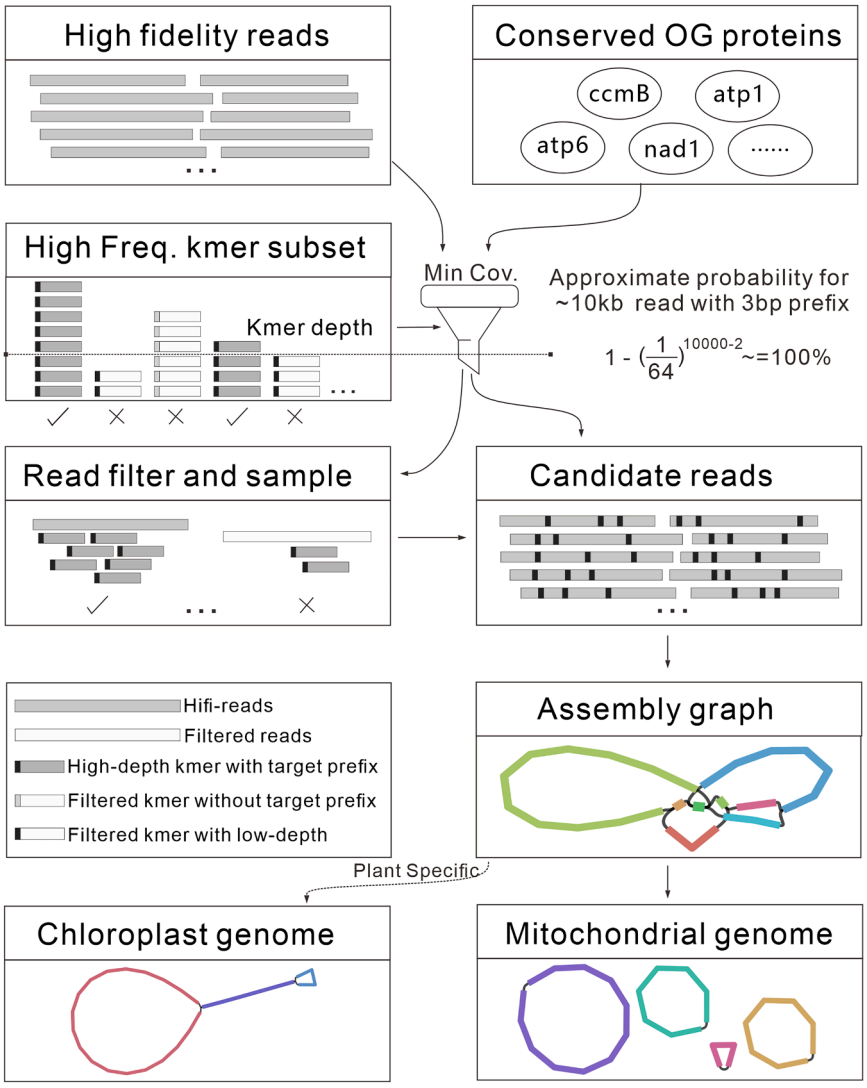


Figure 1. Overview of algorithm optimization and pipeline design in HiMT.

HiFi reads and conserved organelle protein-coding sequences are used to automatically determine an appropriate threshold for high-frequency *k*-mers with a fixed 3-bp prefix. Reads above the high-frequency *k-mer threshold* are identified as candidate sequences for downstream assembly. Once the raw assembly graph is generated, the chloroplast and mitochondrial genomes are separated and assembled individually. OG, organelle genome; Freq, frequency; cov, coverage.

high-frequency *k*-mer table. To further reduce the risk of read loss, HiMT retains four distinct 3-bp prefixes by default, although practical tests suggest that this precaution may not be necessary.

Using the simplified *k*-mer table, HiMT identifies reads containing high-frequency *k*-mers that exceed the depth threshold (Figure 1). This process enables the direct isolation of organelle genome HiFi reads. The selected reads are then used for assembling organelle genomes using circular or graph-based assembly methods, such as Flye (Kolmogorov et al., 2019).

Feature overview of HiMT

We successfully implemented the core operational framework of HiMT and recompiled several dependent software packages to ensure full cross-platform compatibility. HiMT can be deployed natively on Linux, macOS, or Windows systems without reliance on intermediate environments or virtualization programs such as the Windows Subsystem for Linux. The main program includes four key functionalities (Figure 2A): "Assembly" for organelle genome assembly, "Read Filter" for filtering reads without assembling, "Quality Assess" for evaluating assembly results, and "Compare" for comparing different mitochondrial assembly outcomes.

When run in the command-line version, HiMT operates as a fully automated process. Users are only required to provide HiFi sequencing data and define an output directory (Figure 2B). For plant datasets, the output includes eight files, comprising the mitochondrial and chloroplast genomes, as well as interactive assembly quality reports.

For users who prefer graphical interfaces or lack familiarity with command-line tools, HiMT also offers a GUI based on the TBtools-II platform (Figure 2C). The GUI incorporates an intuitive visualization module for the assembly graph, enabling rapid quality assessment of the assembled genomes (Figure 2D).

Evaluation of HiMT assemblies: Completeness and success rate

Assembly completeness and success rate are critical benchmarks for assessing the performance of genome assembly software. In this study, we compared HiMT with PMAT, a state-of-the-art mitochondrial genome assembler. The results show that HiMT achieves mitochondrial genome completeness comparable to that of PMAT. For the pineapple variety 'BL', HiMT assembled a mitochondrial genome of 911 652 bp, identical in length to the circularized genome produced by PMAT (Figure 3A–3C). The two assemblies were nearly identical in contig sequences, with an average nucleotide identity (ANI) of 99.99%. In the case of rice, both tools produced highly consistent results, with assembly lengths of 454 862 bp (HiMT) and 454 865 bp (PMAT) and an ANI of 100% (Figure 3D–3F). Similarly, across 26 accessions from 24 species, HiMT and PMAT produced assemblies of comparable lengths, with ANI values ranging from 99.95% to 100% (Table 1). In most cases, differences between the two tools were negligible and likely due to imperfect read-baiting, extension, or manual de-circularization processes required by PMAT (Supplemental Figure 5).

To evaluate assembly robustness, we tested HiMT and PMAT on 26 species representing diverse plant families using whole-genome HiFi sequencing data. While PMAT successfully

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

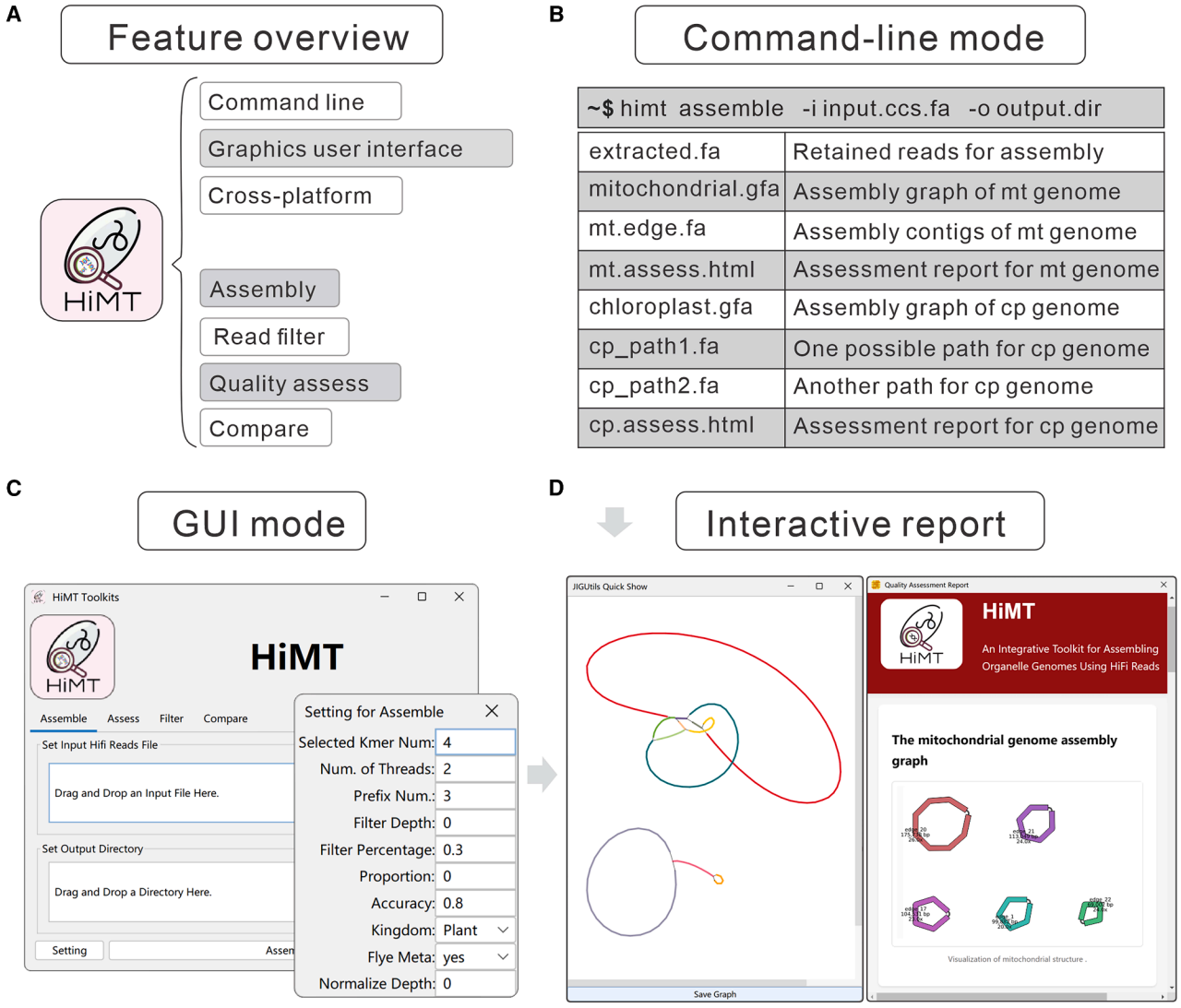


Figure 2. Overview of HiMT features and the GUI.

(A) Three software features and four functional operation modes of HiMT.

(B) Demo of HiMT command-line usage for one-step organelle genome assembly and output files. Mt, mitochondrion; cp, chloroplast.

(C) Demo of the HiMT GUI.

(D) Interactive assembly report demo for the HiMT assembly function.

assembled most organelle genomes based on manually estimated nuclear genome size parameters (Figure 3G; Table 2), there were several instances of assembly failure for either chloroplast or mitochondrial genomes. Some failures could be resolved by adjusting parameters; however, multiple trials were often required. In contrast, HiMT achieved a 93% success rate using default parameters (~100% with fine-tuning parameters), demonstrating both robustness and completeness (Figure 3G; Table 2; Supplemental Figure 1). Unlike PMAT, which frequently required fine-tuning (-fc, subset factor) to handle data reduction, HiMT delivered high-quality assemblies more efficiently with minimal manual intervention. Furthermore, HiMT includes a "Read Filter" function that enables users to integrate HiMT's read selection with external HiFi assemblers such as Canu or PMAT (Koren et al., 2017; Bi et al., 2024), though it primarily uses the built-in Flye. This flexible pairing can produce faster and potentially more accurate results, especially when users do not require cross-platform compatibility or a graphical interface.

Evaluation of HiMT performance: Time and space

A key objective in developing HiMT was to enable reference-quality plant mitochondrial genome assemblies on standard laptops. To evaluate this capability, we compared HiMT, PMAT, and TIPPo software (Xian et al., 2025) across three representative datasets: *Pohlia nutans* (mitochondrial genome ~100 kb), *Ananas comosus* (~1 Mb), and *Musa acuminata* (potentially exceeding 10 Mb) (Figure 4). HiMT produced accurate mitochondrial genome assemblies for all datasets using both default parameters and subsampled data. The resulting genome sizes closely matched those obtained with PMAT after

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

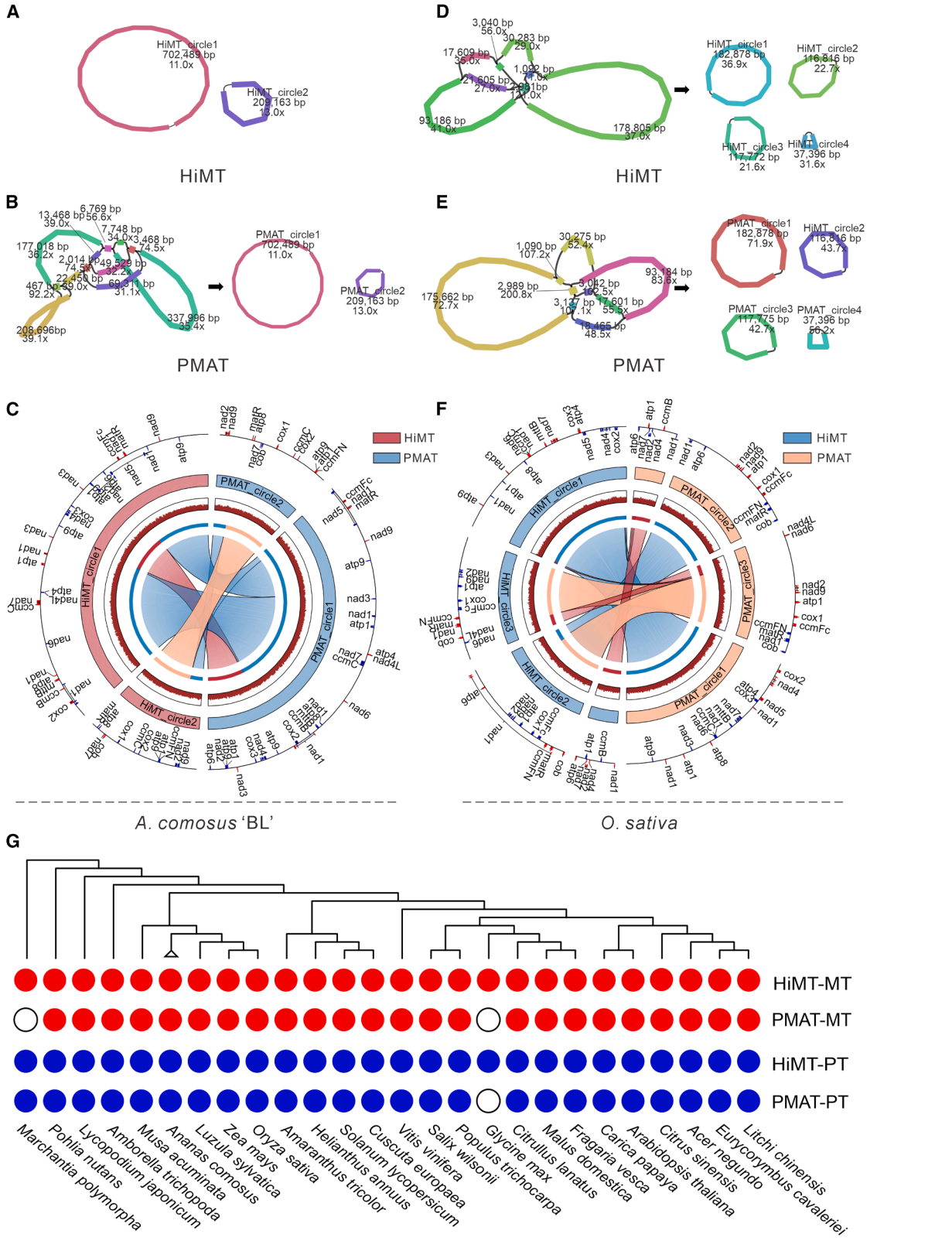


Figure 3. Evaluation of HiMT assembly results.

(A–C) Comparative analysis of the assembly results for the pineapple mitochondrial genome using HiMT and PMAT. **(A)** and **(B)** present assembly visualizations generated via Bandage, while **(C)** features a Circos plot that juxtaposes the two genomes. From the innermost to the outermost rings, the plot illustrates syntenic blocks, GC content, contig names, and the location of conserved genes within the genome.

(legend continued on next page)

extensive parameter tuning and subsampling. In contrast, when run with default parameters, TIPPo and PMAT often failed to complete the assembly or yielded lower-quality results.

HiMT also demonstrated superior runtime performance. It was at least 1.4× faster than PMAT and 2.5× faster than TIPPo, regardless of parameter settings or data subsampling. Memory usage was limited to only 3.40 GB—approximately one-tenth of PMAT's (31.83 GB) and one-twenty-eighth of TIPPo's (97 GB) usage. Additionally, HiMT consumed significantly less disk space than either alternative (Figure 4; Supplemental Table 3).

In summary, HiMT delivers mitochondrial genome assemblies of comparable completeness to those produced by PMAT and TIPPo, while offering substantial advantages in runtime, memory usage, and storage requirements.

Interactive quality assessment report for plant mitochondrial genome assembly results

Although numerous plant mitochondrial genomes have been reported, no current software allows for a rapid evaluation and comparison of mitochondrial genome completeness. To address this limitation, we developed the ''Assess'' mode in HiMT. This feature accepts organelle genome assemblies in GFA or FASTA format and generates a detailed, interactive HTML report summarizing key assembly metrics (Figure 5; Supplemental File 1).

The quality of mitochondrial genome assemblies is typically evaluated using two main criteria: completeness and assembly error rate. Completeness is assessed by detecting conserved mitochondrial genes, whereas the error rate is inferred from abrupt drops in sequencing depth at specific genomic locations.

To facilitate this assessment, HiMT predicts conserved genes in the input sequences and compares their copy number and sequence similarity to reference genomes. These data are presented in an interactive heatmap (Figure 5A). For each contig, HiMT also calculates sequencing depth and sliding-window GC content, displaying these values in an interactive Circos plot (Figure 5B). Additionally, if a previous genome assembly or a closely related species genome assembly is available, HiMT enables rapid comparative analysis by generating a global synteny plot to assess genome-level collinearity and structural similarity between the two assemblies (Figure 3C and 3F).

Support for the assembly and analysis of chloroplast and animal mitochondrial genomes

Plant chloroplast genomes are highly conserved in both sequence and structural organization and are typically sequenced at much greater depth than mitochondrial genomes in green plant tissue datasets. As a result, HiMT achieves near-100% success in assembling chloroplast genomes. HiMT also demonstrates strong performance in assembling animal mitochondrial genomes, consistently generating high-quality results. To date, we have successfully assembled 28 plant chloroplast genomes and 5 animal mitochondrial genomes using HiMT (Supplemental Figures 2–4).

DISCUSSION

A simple Read Filtering process facilitates rapid plant mitochondrial genome assembly

Current software available for assembling plant mitochondrial genomes includes GSAT and PMAT (He et al., 2022; Bi et al., 2024). GSAT constructs a preliminary genome map using short-read data and refines it with long-read data; however, this method struggles with organelle genomes that contain extensive repetitive sequences. PMAT adopts a downsampling approach, selecting a subset of long reads based on a fixed ratio, assembling the dataset, and then extending reads that contain mitochondrial protein-coding genes. This downsampling strategy is not well suited for automation, partly depends on prior knowledge of nuclear genome size, and requires several attempts to optimize parameters. Additionally, downsampling can lead to the loss of certain mitochondrial contigs. During the submission of this manuscript, a new tool, TIPPo, was introduced (Xian et al., 2025). TIPPo applies *k*-mer depth filtering via KMC3 software but consumes excessive memory (>100 GB) for datasets such as *P. nutans* and *A. comosus* and still fails to assemble their mitochondrial genomes (Figure 4). While these tools offer some adjustable filtering parameters, improvements are generally modest and often require multiple trial runs to identify optimal settings.

HiMT addresses these challenges through two simple yet effective optimizations in its filtering algorithm (Figures 1 and 2). First, HiMT estimates genome sequencing depth by reverse-calculating it from the depth of conserved mitochondrial genes. Second, it reduces the computational burden of *k*-mer frequency analysis by restricting the *k*-mer prefix to a limited number of base combinations. These strategies enable HiMT to achieve high assembly success rates and completeness while minimizing computational time and memory usage (Figures 3 and 4). This filtering framework can be extended to other extranuclear genomes or assemblies characterized by uneven sequencing depth. Indeed, HiMT's successful assembly of both chloroplast and animal mitochondrial genomes demonstrates its versatility. Furthermore, by implementing *k*-mer filtering natively rather than relying on external tools, HiMT achieves full cross-platform compatibility.

Strategies for evaluating plant mitochondrial genome assembly and analysis

In nuclear genome research, a variety of metrics, such as CEGMA and BUSCO scores, are commonly used to assess assembly completeness (Parra et al., 2007; Simão et al., 2015). While

(D–F) Comparison of the rice mitochondrial genome assemblies generated by HiMT and PMAT.

(G) Comparative assessment of mitochondrial genome assembly outcomes across diverse species. The phylogenetic tree is sourced from TimeTree. Each row, organized from top to bottom, displays HiMT mitochondrial genome assembly results, PMAT mitochondrial genome assembly results, HiMT chloroplast assembly results, and PMAT chloroplast assembly results. Solid circles indicate successful assemblies, while hollow circles denote assembly failures. MT/CP, mitochondrial/chloroplast genome assemblies.

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

Species	SRA accession	FASTA format data size (GB)	HiMT assembled mitochondrial genome size (bp)	PMAT assembled mitochondrial genome size (bp)	ANI (%)	Repeat content (%)
Pohlia nutans	CRR383826	29 .0	99 734	99 733[a]	99.96	0.86
Cuscuta europaea	ERR9250942	24.2	406 647	406 648[a]	100.00	6.77
Amaranthus tricolor	CRR511440	23.9	382 384	382 432[a]	99.98	3.52
Salix wilsonii	SRR21570388	8.3	711 461	711 456[a]	100.00	2.31
Helianthus annuus	SRR14782853	11.9	300 948	300 887[a]	99.97	12.24
Luzula sylvatica	ERR8705854	4.4	649 009[b]	633 359[a]	99.98	3.05
Lycopodium japonicum	SRR24785435	2.4	454 419	454 458[a]	99.99	24.18
Populus trichocarpa	SRR22064349	6.5	803 754	803 673[a]	99.99	3.01
Malus domestica	ERR6939264	8.8	396 946	396 949[a]	100.00	5.07
Oryza sativa	SRR10238608	23.6	454 862	454 865	100.00	4.61
Ananas comousus 'BL'	SRX22853642	27.3	911 652	911 652	99.99	5.30
Ananas comosus 'LY'	SRX22853646	25.5	928 379	928 401	99.99	5.06
Ananas comosus 'YLL'	SRX22853636	26.1	920 578	920 594	99.99	2.77
Musa acuminata	–	13.3	9 285 210[b]	9 568 932[b]	99.98	25.69
Citrus sinensis	CRR881619	31.2	519 389	519 399	100.00	2.41
Litchi chinensis	–	14.6	858 535	858 532	100.00	2.65
Carica papaya	SRR24464772	9.6	476 890	476 890	100.00	8.23
Arabidopsis thaliana	CRR302668	21.4	367 821[b]	367 809	99.96	16.58
Fragaria vesca	SRR22495382	25.5	312 872	312 870	100.00	2.95
Glycine max	SRR29929794	26.1	614 472[b]	failure	–	45.09
Vitis vinifera	SRR29686483	21.5	816 565	816 565	100.00	4.65
Solanum lycopersicum	SRR15243702	19.6	372 819[b]	371 763[b]	99.95	4.99
Zea mays	SRR15447419	5.3	548 891	521 111	99.99	13.71
Amborella trichopoda	SRR28888927	36.5	3 865 651	3 746 734 (not circle)	99.97	20.54
Marchantia polymorpha	SRR26200312	6.2	186 295	failure	–	10.94
Citrullus lanatus	SRR26555778	22.1	379 228	379 235	100	3.78
Acer negundo	–	8.6	579 640	579 664	100	3.36
Eurycorymbus cavaleriei	–	17.5	562 195	562 212 (not circle)	99.99	1.96

Table 2. Comparison of mitochondrial assemblies generated by HiMT and PMAT

For datasets lacking accession numbers and not available in public repositories, please contact the corresponding author to request access.

[a]Data sourced from the PMAT publication.

[b]Sequences for which circularization was not resolved.

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

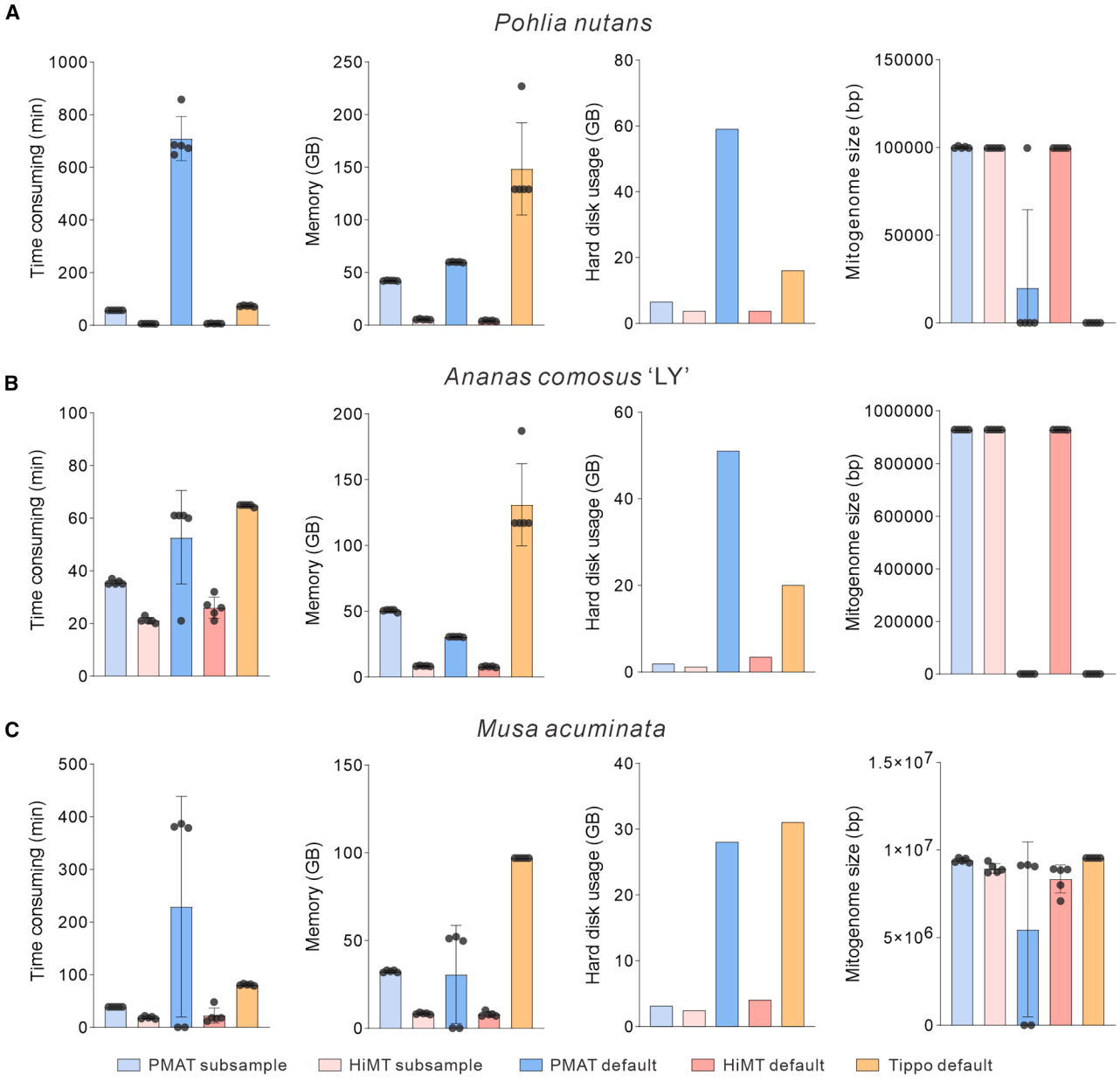


Figure 4. Comparison of assembly efficiency across HiMT, PMAT, and TIPPo.

(A) Computational resources required for assembling the *P. nutans* mitochondrial genome are displayed from left to right: runtime, maximum memory usage, disk space usage, and final mitochondrial genome size. Each solid circle represents one of five replicate runs. Hollow arrows indicate the performance of HiMT.

(B) Computational resource usage for assembling the *A. comosus* mitochondrial genome.

(C) Computational resource usage for assembling the *M. acuminata* mitochondrial genome.

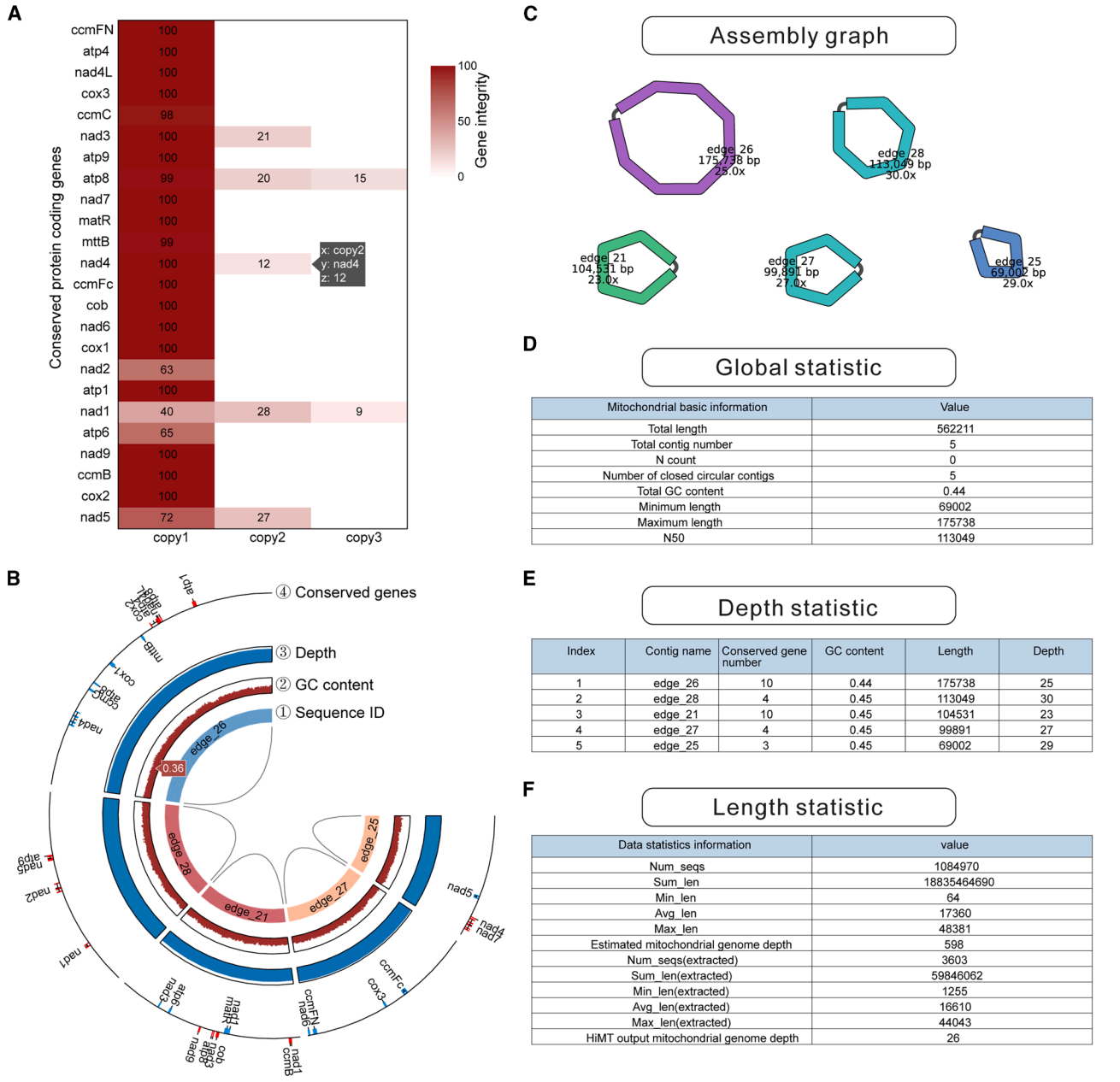
several strategies are implicitly or repeatedly applied in the literature, no standardized method or consensus exists for evaluating the quality of plant mitochondrial genome assemblies. A major contributing factor to this gap is the absence of dedicated software tools specifically designed for mitochondrial genome evaluation.

To address this gap, HiMT integrates widely used assessment strategies into a unified workflow, generating an interactive mitochondrial genome quality report in one step. This feature supports evaluation not only of assemblies generated by HiMT but also of publicly available mitochondrial genomes or those produced by alternative assembly software.

Efficient assembly strategies for complex mitochondrial genomes and conformational insights

We believe that HiMT currently offers the most effective one-click solution for the automated assembly of plant mitochondrial genomes. However, a small number of species with highly complex mitochondrial genomes—such

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467



Mitochondrial basic information	Value
Total length	562211
Total contig number	5
N count	0
Number of closed circular contigs	5
Total GC content	0.44
Minimum length	69002
Maximum length	175738
N50	113049

Index	Contig name	Conserved gene number	GC content	Length	Depth
1	edge_26	10	0.44	175738	25
2	edge_28	4	0.45	113049	30
3	edge_21	10	0.45	104531	23
4	edge_27	4	0.45	99891	27
5	edge_25	3	0.45	69002	29

Data statistics information	value
Num_seqs	1084970
Sum_len	18835464690
Min_len	64
Avg_len	17360
Max_len	48381
Estimated mitochondrial genome depth	598
Num_seqs(extracted)	3603
Sum_len(extracted)	59846062
Min_len(extracted)	1255
Avg_len(extracted)	16610
Max_len(extracted)	44043
HiMT output mitochondrial genome depth	26

Figure 5. Assessment report for the completeness of plant mitochondrial genome assembly.

(A) Interactive heatmap illustrating the sequence similarity of conserved genes in the mitochondrial assembly. Darker colors represent higher similarity. The vertical axis lists the IDs of the conserved genes, while the horizontal axis corresponds to different gene copies.

(B) Overview plot of mitochondrial genome sequencing and rapid annotation. The internal lines indicate circular sequences. The rings, from innermost to outermost, represent the contig IDs, sliding-window GC content, contig depth, and the positions of conserved genes.

(C) Visualization of the organelle genome assembly.

(D) Length distribution of the mitochondrial genome assembly contigs.

(E) Number of annotated genes, GC content, contig length, and average sequencing depth per contig.

(F) Statistical information in input and HiMT-filtered data, including estimated mitochondrial genome depth.

as banana, whose mitochondrial genome may exceed 10 Mb —remain particularly challenging. To date, no available tools, including PMAT, HiMT, or TIPPo, have successfully generated high-quality mitochondrial genome assemblies for these species. Even with the integration of Nanopore ultra-long reads, our attempts did not yield satisfactory results. Improving the assembly accuracy of such complex genomes will require further optimization, refinement of algorithms, and iterative parameter adjustment across multiple tools, including HiMT.

Recent research has increasingly highlighted the dynamic conformational structures of plant mitochondrial genomes. However, current assembly software—including HiMT—has limited

Please cite this article in press as: Tang et al., HiMT: An integrative toolkit for assembling organelle genomes using HiFi reads, Plant Communications (2025), https://doi.org/10.1016/j.xplc.2025.101467

capacity to resolve and represent these structural variations. Moving forward, a key priority for HiMT development will be enhancing its capacity to capture mitochondrial conformational diversity, potentially through the integration of more sophisticated algorithms or modules for structural resolution.

Conclusion

By applying a simple yet effective strategy based on sequencing depth estimation and streamlined *k*-mer frequency analysis, we have addressed key challenges in plant mitochondrial genome assembly. We have improved the methodological efficiency, reduced computational time, and minimized resource consumption. Furthermore, HiMT software is user-friendly, runs across multiple platforms on standard laptops, and features a GUI that enables users to perform organelle genome assembly. It also generates interactive, comprehensive assembly evaluation reports. We are confident that HiMT will have broad applicability in cellular organelle research, with particular value for advancing plant mitochondrial genome research.

METHODS

Implementation of the read filtering method

Data reduction

WGS datasets are typically large, and directly processing the full volume of raw reads can hinder filtering efficiency and significantly increase time and storage demands during assembly with Flye (Kolmogorov et al., 2019). HiMT addresses this challenge by applying a random, uniform sampling strategy based on file size. For instance, if the input data is 100 GB, HiMT samples 10% (10 GB) for downstream analysis. Given the typically high mitochondrial genome coverage in WGS datasets, this sampling step effectively reduces data redundancy without compromising the quality of mitochondrial genome assembly. If the resulting mitochondrial genome depth falls below a predefined threshold, HiMT automatically increases the sampling rate to retain more data.

Determination of filtering thresholds

To estimate sequencing depth and guide filtering, HiMT uses amino acid sequences from 24 conserved mitochondrial protein-coding genes obtained from published mitochondrial genomes of *Arabidopsis thaliana* (NC_037304) and *Oryza sativa* (NC_066488) (Sloan et al., 2018; Jiang et al., 2022), as well as sequences from *Drosophila gunungcola* (CM045947) for animal mitochondrial genome analysis. Using tBLASTn, HiMT identifies reads aligning to each conserved gene to estimate gene-level sequencing depth. Due to sequence duplications and nuclear-mitochondrial transfers, these depths vary significantly. The minimum sequencing depth among all conserved genes is used as a rough estimate of the overall minimum sequencing depth for the lowest genome-wide mitochondrial coverage. To account for strand complementarity and depth fluctuations, HiMT applies a default scaling factor of 0.3 to this minimum depth to define a threshold for distinguishing high-frequency from low-frequency *k*-mers.

Acquisition of high-frequency k-mer set

By default, HiMT divides reads into 21-mers and calculates the frequency of each *k*-mer. Instead of storing all *k*-mers and their counts—which requires substantial memory—HiMT selectively retains a subset using a prefix-based categorization strategy. For instance, fixing the first three bases of a *k*-mer yields 64 possible categories ($4 \times 4 \times 4$). HiMT retains only 4 of these groups by default, reducing memory usage to one-sixteenth of that required for storing the complete *k-mer* set.

Read extraction

The *k-mer count* is analyzed for each read, and if more than 80% of the *k*-mers are classified as high frequency, the read is retained for mitochondrial genome assembly. By default, HiMT limits the retained reads to a mitochondrial genome depth of no more than 50× to optimize Flye assembly performance. However, users can adjust parameters to analyze the original depth or specify a custom target depth for the assembled mitochondrial genome.

Implementation of the assembly function

Once high-depth sequencing reads are extracted, HiMT performs organelle genome assembly using Flye, an open-source assembler optimized for long-read data and originally designed for Linux operating systems. To enable full cross-platform compatibility, we recompiled Flye for use on Windows systems. By default, HiMT operates in Flye's meta mode, which allows for the simultaneous assembly of both chloroplast and mitochondrial genomes.

Implementation of the quality assessment report

HiMT implements the Plotly library in Python, a powerful open-source data visualization tool well-suited for generating dynamic and interactive charts. Users can input a GFA or FASTA file to receive a user-friendly HTML format genomic quality assessment report.

Heatmap generation

The conserved mitochondrial gene set is mapped to the assembled genome using miniprot (Li, 2023). The resulting alignments are visualized as a heatmap using Plotly's built-in functionality, enabling rapid assessment of gene presence and sequence similarity across contigs.

Circos plot generation

The Circos plot implemented in HiMT is developed using Plotly's foundational plotting tools. Based on BLASTn alignment results, the plot highlights conserved gene duplications and large syntenic blocks within the mitochondrial genome. Repetitive elements are connected via Bezier curves. A sliding-window approach is applied to calculate GC content across each mitochondrial genome. Additionally, sequencing depth and conserved gene locations for each contig are integrated, resulting in a multi-layered visual representation of genome features.

Table generation

HiMT summary tables provide key statistics on filtered input data, including data volume and estimated mitochondrial genome depth. They also report core assembly metrics such as total genome size, GC content, and N50 value. These tables are embedded in the quality report to support a comprehensive evaluation.

Comparison of various software for assembling organelle genomes

HiMT employs default parameters to assemble mitochondrial genomes from evolutionarily diverse species. For PMAT, we utilized the built-in -fc (subset factor) parameter to obtain a successful assembly process with data downsampling. For an evaluation of computational resources, we tested the resource consumption of assembly under both the default settings and with manually adjusted subsample parameters. Additionally, we assessed the resource consumption of TIPPo under default conditions. All tests were conducted sequentially on the same server, which features an Intel Xeon Platinum 8488C CPU processor. The specific assembly commands can be found in Supplemental Tables 1–3.

ANI calculation and repetitive sequence identification

The ANI between sequences was calculated using FastANI (Jain et al., 2018). Repetitive sequences were identified using BLASTn (Altschul et al., 1990), with the "-evalue 1e-5" parameter added to the default settings. The detected repetitive sequences were merged to avoid redundant counting of the same regions.

DATA AVAILABILITY

All datasets, along with accession numbers, are publicly available (Supplemental Tables 1 and 2). The assemblies generated and analyzed during this study are available on GitHub (https://github.com/tang-shuyuan/data-for-HiMT).

HiMT is freely available to non-commercial users via GitHub (https://github.com/tang-shuyuan/HiMT), while its GUI version is available through the TBtools Plugin Store.

FUNDING

This research was supported by the Science and Technology Special Fund of Hainan Province (ZDYF2025XDNY106), the Project of State Key Laboratory of Tropical Crop Breeding (SKLTCBZRJJ202502) and the National Natural Science Foundation of China (32102320). C.C. and A.L. were supported by the Central Public-interest Scientific Institution Basal Research Fund for the Chinese Academy of Tropical Agricultural Sciences (1630032024026 and 1630032024001).

ACKNOWLEDGMENTS

We thank all members of the Xia lab, He lab, and Xu lab for their generous help. We extend our appreciation to Guiyang Watchbio Co., Ltd. No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

C.C. conceived the project. C.C. and S.T. designed the toolkit functions. S.T. implemented all Python coding, while C.C. developed the HiMT plugin for TBtools-II. Y. Liang and F.W. tested all functions and refined the figures. Y. Wu, J.L., L. Lin, J.F., A.L., and Y. Wang tested the functions and assisted in preparing the tutorial manual. C.X., J.W., Y.Z., and R.X. provided valuable feedback during the development of HiMT. C.C. and S.T. prepared most of the figures and wrote the manuscript. All authors read and approved the final version of the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: April 17, 2025
Revised: June 8, 2025
Accepted: July 24, 2025

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A. D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19**:455–477. https://doi.org/10.1089/cmb.2012.0021.

Bi, C., Shen, F., Han, F., Qu, Y., Hou, J., Xu, K., Xu, L.A., He, W., Wu, Z., and Yin, T. (2024). PMAT: an efficient plant mitogenome assembly toolkit using low-coverage hifi sequencing data. Hortic. Res. **11**: uhae023. https://doi.org/10.1093/hr/uhae023.

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y., Feng, J., Chen, H., He, Y., et al. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol. Plant **16**:1733–1742. https://doi.org/10.1016/j.molp.2023.09.010.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18**:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. **17**:134. https://doi.org/10.1186/s13059-016-1004-2.

He, W., Xiang, K., Chen, C., Wang, J., and Wu, Z. (2023). Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Briefings Bioinf. **24**:bbac522. https://doi.org/10.1093/bib/bbac522.

Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A.O., Liang, F., Li, J., Sandoval, J.R., Cooper, D.N., Ye, K., et al. (2024). NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. **25**:107. https://doi.org/10.1186/s13059-024-03252-4.

Jiang, H., Lu, Q., Qiu, S., Yu, H., Wang, Z., Yu, Z., Lu, Y., Wang, L., Xia, F., Wu, Y., et al. (2022). Fujian cytoplasmic male sterility and the fertility restorer gene osrf19 provide a promising breeding system for hybrid rice. Proc. Natl. Acad. Sci. USA **119**:e2208759119. https://doi.org/10.1073/pnas.2208759119.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. **9**:5114. https://doi.org/10.1038/s41467-018-07641-9.

Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., and Li, D.Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. **21**:241. https://doi.org/10.1186/s13059-020-02154-5.

Kang, B.-C., Bae, S.-J., Lee, S., Lee, J.S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., and Kim, J.-S. (2021). Chloroplast and mitochondrial dna editing in plants. Nat. Plants **7**:899–905. https://doi.org/10.1038/s41477-021-00943-9.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. **37**:540–546. https://doi.org/10.1038/s41587-019-0072-8.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736. https://doi.org/10.1101/gr.215087.116.

Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics **39**:btad014. https://doi.org/10.1093/bioinformatics/btad014.

Martin, W., Deusch, O., Stawski, N., Grünheit, N., and Goremykin, V. (2005). Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. Trends Plant Sci. **10**:203–209. https://doi.org/10.1016/j.tplants.2005.03.007.

Mower, J.P., Sloan, D.B., and Alverson, A.J. (2012). Plant mitochondrial genome diversity: the genomics revolution. In J.F. Wendel, J. Greilhuber, J. Dolezel, and I.J. Leitch, eds., *Volume 1*, Plant Genome Diversity (Vienna: Springer Vienna), pp. 123–144. https://doi.org/10.1007/978-3-7091-1130-7_9.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23**:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., and Vendramin, G.G. (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol. Ecol. **14**:689–701. https://doi.org/10.1111/j.1365-294x.2004.02410.x.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Sloan, D.B., Wu, Z., and Sharbrough, J. (2018). Correction of persistent errors in arabidopsis reference mitochondrial genomes. Plant Cell **30**:525–527. https://doi.org/10.1105/tpc.18.00024.

Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. Trends Plant Sci. **27**:391–401. https://doi.org/10.1016/j.tplants.2021.10.006.

Wang, J., Kan, S., Liao, X., Zhou, J., Tembrock, L.R., Daniell, H., Jin, S., and Wu, Z. (2024). Plant organellar genomes: much done, much more to do. Trends Plant Sci. **29**:754–769. https://doi.org/10.1016/j.tplants.2023.12.014.

Wynn, E.L., and Christensen, A.C. (2019). Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda) **9**:549–559. https://doi.org/10.1534/g3.118.200948.

Xian, W., Bezrukov, I., Bao, Z., Vorbrugg, S., Gautam, A., and Weigel, D. (2025). TIPPo: a user-friendly tool for de novo assembly of organellar genomes with high-fidelity data. Mol. Biol. Evol. **42**:msae247. https://doi.org/10.1093/molbev/msae247.