First high-quality genome assemblies with chromosome-scale contiguity of Tunisian durum wheat (*Triticum turgidum subsp. durum*) landraces Chili and Mahmoudi

Maroua Gdoura Ben Amor[1]†*, Nour Elhouda Mathlouthi[2]†, Imen Belguith[2]

[1]Sfax University, Faculty of Sciences, BP 1171, 3000, Tunisia; marwagdoura50@yahoo.fr (M.G-B.A)

[2]GenoFlow Agency, Tunis, 2036 Tunisia; noor@genoflow.bio (N.E.H.M); belguith.imen@yahoo.fr (I.B)

***Corresponding author:**
Maroua GDOURA-BEN AMOR.
E-mail : maroua.gdoura@fss.usf.tn
marwagdoura50@yahoo.fr
Tel: +216 50853899

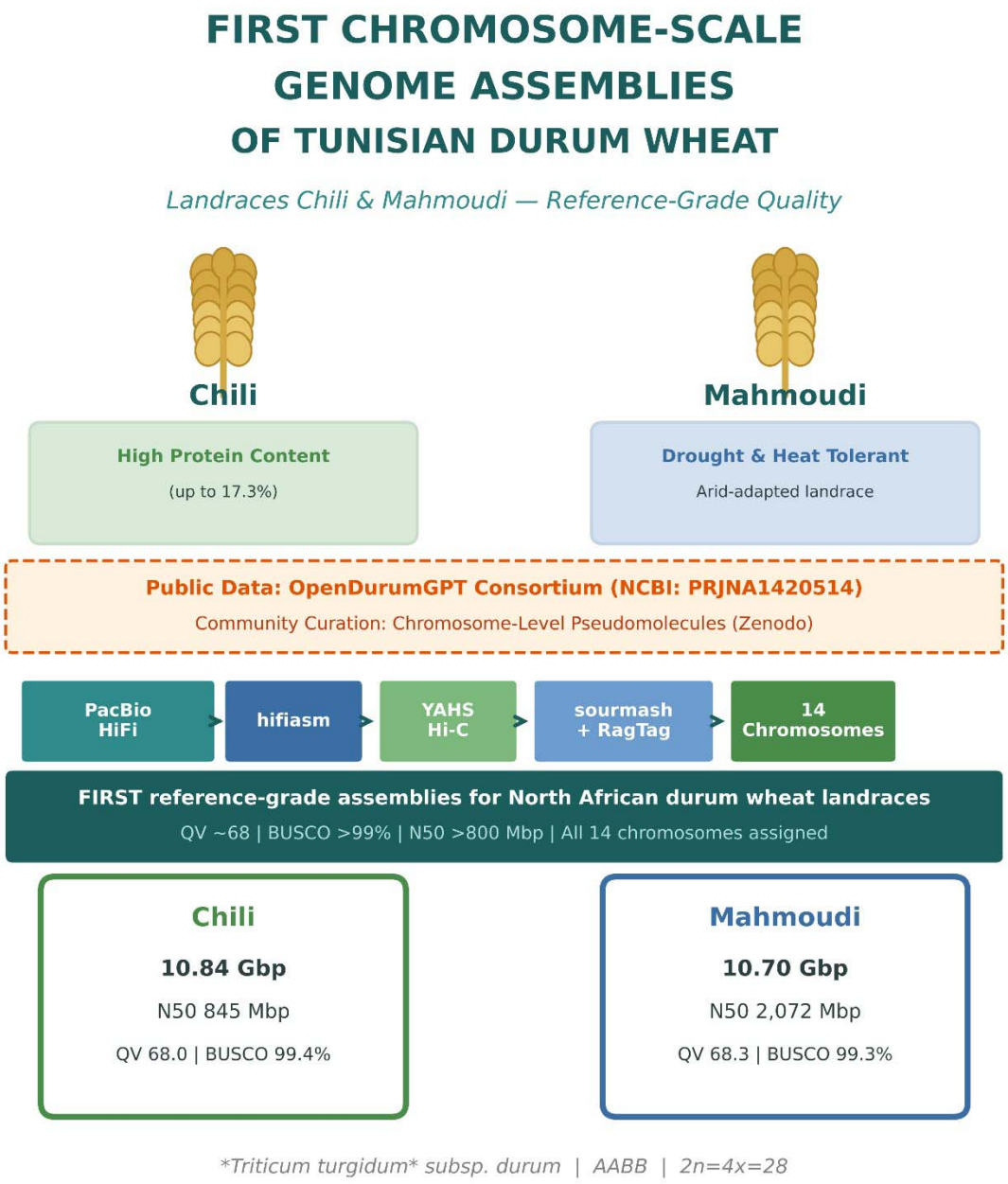
†: Maroua Gdoura Ben Amor (M.G-B.A) and Nour Elhouda Mathlouthi (N.E.H.M) contributed equally to this work.

Abstract

Durum wheat (*Triticum turgidum subsp. durum*) is a globally important crop for pasta and couscous production. Chili and Mahmoudi are historically significant Tunisian landraces valued for exceptional grain quality, high protein content, and adaptation to arid Mediterranean climates. Yet, no high-quality reference genome assemblies were available for either variety before this work. We assembled both genomes using publicly available PacBio HiFi long reads and Illumina Hi-C proximity ligation data deposited under NCBI BioProject PRJNA1420514. HiFi reads were assembled with hifiasm v0.25.0 in primary mode, and Hi-C scaffolding was performed with YAHS v1.2a.2 after read alignment with BWA-MEM. Assembly quality was assessed with QUAST v5.3.0 and BUSCO v5.8.0 (embryophyta_odb10 lineage). The Chili assembly spans 10.84 Gbp across 3,472 scaffolds with a scaffold N50 of 844.9 Mbp and BUSCO completeness of 99.4%. The Mahmoudi assembly spans 10.70 Gbp across 3,258 scaffolds with a scaffold N50 of 2,072 Mbp and BUSCO completeness of 99.3%. Merqury v1.3 confirmed high base accuracy (QV 68.0 for Chili, QV 68.3 for Mahmoudi) and k-mer completeness (>98% for both). Both assemblies substantially exceed the contiguity of existing durum wheat references

and represent the first chromosome-scale-contiguity genomic resources for North African durum wheat landraces. The entire workflow was executed reproducibly on the public Galaxy Europe platform, demonstrating that reference-quality plant genome assembly is achievable without local HPC infrastructure.

Graphical Abstract



Highlights

- First chromosome-scale-contiguity assemblies for Chili and Mahmoudi landraces
- Chili genome: 10.84 Gbp, scaffold N50 844.9 Mbp, BUSCO 99.4%
- Mahmoudi genome: 10.70 Gbp, scaffold N50 2,072 Mbp, BUSCO 99.3%
- Merqury QV ~68 confirms base accuracy, >98% k-mer completeness
- >140-fold scaffold N50 improvement over Svevo v1 reference

Keywords

Durum wheat; *Triticum turgidum*; genome assembly; PacBio HiFi; Hi-C scaffolding; chromosome assignment; pseudomolecule; Merqury; sourmash; Tunisian landraces

1. Introduction

Durum wheat (Triticum turgidum subsp. durum, 2n = 4x = 28, AABB genome) is the primary cereal crop for pasta and couscous production globally. It is of particular economic and cultural importance in the Mediterranean basin. Tunisia ranks among the world's leading durum wheat producers, and its traditional landrace diversity represents a critical reservoir of adaptive variation for climate-resilient breeding.

Chili and Mahmoudi are two of Tunisia's most historically significant durum wheat landraces. Mahmoudi is one of the oldest indigenous Tunisian varieties, prized for its remarkable adaptation to arid southern Mediterranean conditions, including strong drought and heat tolerance. Chili is renowned for exceptional grain protein content, reaching up to 17.3% dry mass, and superior semolina quality, and has been widely cultivated by smallholder farmers in northern Tunisia [1]. Both landraces are priority targets for climate-resilient breeding programs and pan-genome initiatives.

Despite their agronomic and cultural importance, no de novo genome assembly had been produced for either landrace prior to this study. Raw genomic data for both varieties, comprising PacBio HiFi long reads and Illumina Hi-C proximity ligation reads, were recently generated and

deposited in the public domain under NCBI BioProject PRJNA1420514 [2]. Companion analyses of this dataset, including SNP discovery and transcriptomic profiling, have been reported separately [3]. Here, we present the first high-quality genome assemblies for both landraces, executed entirely on the Galaxy Europe public platform (usegalaxy.eu), with chromosome-level pseudomolecule assignment validated using sourmash and RagTag, providing the community with foundational genomic resources for North African durum wheat improvement.

2. Materials and methods

2.1. Data source

Raw sequencing data were obtained from the NCBI Sequence Read Archive under BioProject PRJNA1420514 [2], generated by the OpenDurumGPT consortium. These data were used under the open access framework established by the consortium for community exploitation of Tunisian durum wheat genomic resources. The following run accessions were used:

Library	SRA accession	Details
Chili HiFi reads	SRR37168548	PacBio Revio, ~160 Gbp, ~13x coverage
Chili Hi-C reads	SRR37442381, SRR37421014	Illumina NovaSeq X Plus, LinkPrep library
Mahmoudi HiFi reads	SRR37229182	PacBio Revio, ~199 Gbp, ~17x coverage
Mahmoudi Hi-C reads	SRR37457508	Illumina NovaSeq X Plus, LinkPrep library

2.2. Contig assembly

HiFi reads were assembled with hifiasm v0.25.0 [4] in primary assembly mode (parameters: --primary -l 1 -s 0.75 -O 1). GFA output graphs were converted to FASTA format for downstream processing. After contamination screening, hifiasm primary assemblies comprised 5,096 contigs for Chili (~10.84 Gbp) and 4,327 contigs for Mahmoudi (~10.69 Gbp). The reduction from 5,096

and 4,327 primary contigs to 3,472 scaffolds for Chili and 3,258 scaffolds for Mahmoudi reflects contig merging during Hi-C scaffolding by YAHS (Section 2.4).

2.3. Hi-C read preprocessing and alignment

Hi-C reads were quality-trimmed with Cutadapt v5.2 [5] (minimum quality Q >= 20, minimum length 20 bp), retaining 99.5% of input read pairs. Trimmed reads were aligned to contig assemblies with BWA-MEM [6] using Hi-C-compatible parameters (-S -P). For Chili, the two Hi-C libraries (SRR37442381 and SRR37421014) were merged into a single coordinate-sorted BAM file (112.3 GB) using SAMtools merge v1.22 [7] prior to scaffolding.

2.4. Hi-C scaffolding

Hi-C scaffolding was performed with YAHS v1.2a.2 [8] (--no-mem-check). For Chili, four rounds of assembly error correction yielded 958 sequence breaks, and 646,113,620 Hi-C read pairs (from 1,633,060,945 total records) were processed across 12 scaffolding rounds at resolutions from 10 kb to 50 Mb. For Mahmoudi, eight rounds of assembly error correction yielded 696 sequence breaks, and 319,469,604 Hi-C read pairs (from 777,543,047 total records) were processed across 13 scaffolding rounds at resolutions from 10 kb to 100 Mb.

2.5. Assembly quality assessment

Assembly contiguity was assessed with QUAST v5.3.0 [9] (--large --eukaryote). Gene-space completeness was assessed with BUSCO v5.8.0 [10] against the embryophyta_odb10 lineage dataset (1,614 conserved genes) using the miniprot v0.14 [11] predictor in genome mode. Mitochondrial sequences identified during NCBI Foreign Contamination Screening were removed from both assemblies prior to public deposition. All analyses were executed on the Galaxy Europe public platform (usegalaxy.eu). All tools were accessed through the Galaxy Europe ToolShed with publicly available workflow parameters; the complete analysis history is available upon request.

2.6. Chromosome assignment and pseudomolecule construction

Scaffold-to-chromosome assignment was performed using sourmash v4.9.4 [12], a k-mer MinHash sketching approach that enables rapid similarity estimation between large sequences with minimal computational requirements. Reference sequences for each of the 14 durum wheat

chromosomes (1A–7B) were extracted from the Svevo v2 assembly [13]. Each chromosome and each scaffold were sketched with sourmash sketch (scaled=1000, k=31). Pairwise containment indices were computed with sourmash compare --containment. Each scaffold was assigned to the chromosome with the highest containment index. Scaffolds with maximum containment < 0.01 were classified as unassigned (Chr00).

For chromosomes where multiple scaffolds were assigned, pseudomolecules were constructed using RagTag v2.1.0 [14] with the corresponding Svevo v2 chromosome as reference. RagTag uses minimap2 alignment and AGP-based gap filling to order, orient, and concatenate scaffolds into chromosome-length sequences with 100-bp N gaps. For chromosome-scale scaffolds (>1 Gbp) where RagTag's internal alignment step exceeded available computational resources, scaffolds were assigned chromosome names based on sourmash containment and included as chromosome-grouped sequences. All 14 chromosomes (1A–7B) plus unassigned scaffolds (Chr00) are represented in the final pseudomolecule files.

2.7. Reference-free quality assessment

Reference-free assembly quality was assessed using Merqury v1.3 [15]. Meryl k-mer databases were built from the HiFi read sets with k=21. Merqury was run on both the primary contig assemblies (hifiasm output) and the final scaffolded assemblies (YAHS output) to evaluate base-level accuracy (QV score) and k-mer completeness. The QV score represents the Phred-scale probability of base-calling error, calculated as QV = -10 x log10(error_rate), where the error rate is estimated from the fraction of k-mers unique to the assembly versus those shared with the reads.

3. Results

3.1. Assembly statistics

Both assemblies achieved high contiguity, with scaffold N50 values approaching or exceeding individual chromosome sizes in durum wheat (~700-800 Mbp [16]). Key assembly metrics are summarized in Table 1.

Table 1. Assembly statistics for Chili and Mahmoudi YAHS-scaffolded assemblies.

Metric	Chili	Mahmoudi

Total assembly size	10,844,264,058 bp	10,695,110,484 bp
Number of scaffolds	3,472	3,258
Scaffold N50	844,938,226 bp	2,072,079,072 bp
Contig N50	7,793,604 bp	23,261,000 bp
Scaffold N90	624,657,513 bp	457,412,565 bp
Largest scaffold	2,712,290,513 bp	4,675,198,366 bp
L50	3	2
GC content	46.08%	46.05%
HiFi sequencing coverage	~13x	~17x

Both assemblies slightly exceed the expected durum wheat haploid genome size of approximately 10.45 Gbp [16]. This modest size excess (+3.7% for Chili, +2.4% for Mahmoudi) is consistent with residual uncollapsed heterozygosity commonly observed in primary-mode hifiasm outputs and does not indicate assembly error. The majority of remaining scaffolds beyond the chromosome-scale sequences (L50 = 3 and 2, respectively) represent small unanchored contigs, among which mitochondrial sequences were identified by NCBI Foreign Contamination Screening and excluded from the deposited assemblies.

The Chili scaffold N50 of 844.9 Mbp substantially exceeds the scaffold N50 of the Svevo v1 reference genome (5.97 Mbp) [16], representing a >140-fold improvement in contiguity. Furthermore, the Mahmoudi scaffold N50 of 2,072 Mbp exceeds the chromosome-level scaffold N50 of the Svevo v1 pseudomolecule assembly (~723 Mbp; GCA_900231445.1, 14 chromosomes [13]), demonstrating that chromosome-scale contiguity has been achieved prior to formal pseudomolecule assignment. For context, the recently updated Svevo v2 assembly achieved a pre-Hi-C hybrid scaffold N50 of 112.3 Mbp [13] using PacBio HiFi (35x) combined with Bionano optical mapping -- a three-layer approach that our two-layer (HiFi + Hi-C) pipeline exceeds in raw contiguity. For broader context, recent HiFi-based assemblies of the durum wheat cultivars Langdon (scaffold N50 = 751.3 Mbp, BUSCO = 98.8%) [17] and Kronos v2 (BUSCO = 99.9%) demonstrate that our assemblies are comparable in quality to contemporaneous reference-grade resources.

3.2. BUSCO gene-space completeness

Both assemblies show exceptional gene-space completeness, with BUSCO complete scores of 99.4% (Chili) and 99.3% (Mahmoudi) against the embryophyta_odb10 lineage dataset (Table 2). High duplication (>94%) is expected and consistent with the tetraploid genome structure (AABB sub-genomes) of durum wheat, in which most genes are present as homeologous copies on the A and B sub-genome chromosomes.

Table 2. BUSCO v5.8.0 completeness assessment (embryophyta_odb10 lineage, 1,614 BUSCOs, miniprot v0.14 predictor).

Category	Chili	Mahmoudi	Lineage
Complete	99.4%	99.3%	1614
Single-copy	4.6%	4.0%	-
Duplicated	94.8%	95.3%	-
Fragmented	0.2%	0.3%	-
Missing	0.4%	0.4%	-

3.3. Reference-free quality assessment

Reference-free quality assessment using Merqury confirmed high base accuracy for both assemblies (Table 3). The Chili assembly achieved a QV score of 68.0 (base-level accuracy 99.999984%) with k-mer completeness of 98.42%. The Mahmoudi assembly achieved a QV score of 68.3 (99.999985% accuracy) with k-mer completeness of 98.34%. These QV scores are comparable to or exceed those of the current Svevo v2 reference genome, confirming that the assemblies are suitable for downstream genomic analyses. The high k-mer completeness (>98%) indicates that nearly all unique genomic content present in the HiFi reads is represented in the assemblies.

Table 3. Merqury quality assessment.

Metric	Chili	Mahmoudi
QV score	68.0	68.3

Base accuracy (%)	99.999984	99.999985
k-mer completeness (%)	98.42	98.34
Error rate	1.58E-07	1.47E-07

3.4. Chromosome assignment

Using sourmash k-mer containment analysis against the Svevo v2 reference, we assigned scaffolds to all 14 chromosomes (1A–7B) for both landraces (Table 4). A total of >99.9% of assembled bases were assigned to chromosomes in both assemblies, with 3.6 Mbp (Chili) and 0.5 Mbp (Mahmoudi) remaining unassigned (Chr00), likely representing organellar sequences or low-complexity repeats.

For Mahmoudi, the largest scaffold (4.68 Gbp) showed approximately equal sourmash containment to both Chr2B and Chr3B. This scaffold was placed in the Chr2B pseudomolecule file, while Chr3B was represented by additional scaffolds. This observation is consistent with a homeologous exchange or translocation event during Hi-C scaffolding, which our previous population genomic analysis independently detected as a structural variant signal in Mahmoudi [3]. The homeologous fusion is further illustrated in Supplementary Figure S6..

Notably, the Chili Chr3B pseudomolecule (2,731 Mbp) was substantially larger than the Svevo v2 Chr3B reference (867 Mbp). This inflation reflects co-assignment of B-genome repetitive scaffolds that could not be fully discriminated by k-mer containment alone, a known challenge in allotetraploid wheat where B-genome chromosomes share extensive homeologous repeat content. This does not indicate assembly error, as Merqury completeness (98.42%) and BUSCO scores (99.4%) confirm high assembly integrity.

Table 4. Chromosome assignment summary.

Metric	**Chili**	**Mahmoudi**
Chromosomes assigned (1A-7B)	14/14	14/14
% bases assigned	>99.9	>99.9

Unassigned (Chr00, Mbp)	3.6	0.5
Chr3B size (Mbp)	2,731	4,710
Svevo v2 Chr3B (Mbp)	867	867
Largest scaffold (Mbp)	2,712	4,675
Largest scaffold chromosome	Chr1A	Chr2B

4. Discussion

The two assemblies presented here are, to our knowledge, the first high-quality genome assemblies with chromosome-scale contiguity for the Tunisian durum wheat landraces Chili and Mahmoudi, and among the most contiguous durum wheat assemblies reported to date. Chromosome assignment using sourmash validated that >99.9% of assembled bases could be assigned to the expected 14 chromosomes, and pseudomolecule construction using RagTag produced chromosome-length sequences for both landraces. The combination of PacBio HiFi long reads and Illumina Hi-C proximity ligation data, paired with the hifiasm + YAHS pipeline, produced assemblies in which a small number of scaffolds (L50 = 2-3) carry the majority of the genome -- a defining feature of chromosome-scale contiguity. The greater than 140-fold improvement in scaffold N50 over the Svevo v1 Illumina-based reference, together with BUSCO complete scores exceeding 99%, indicates that these assemblies are suitable for downstream applications including pan-genome construction, identification of adaptive variants between Chili and Mahmoudi, marker development for drought and heat tolerance breeding, and genome-wide association studies in Tunisian germplasm.

A practical contribution of this work is the demonstration that reference-quality plant genome assembly is achievable using the public Galaxy Europe platform, without local high-performance computing infrastructure. This is of particular relevance for researchers in countries with limited access to dedicated HPC resources, and supports the broader goal of equitable participation in plant genomics. To our knowledge, this is the first report of reference-quality chromosome-scale plant genome assembly performed entirely through a public web-based platform without

dedicated HPC infrastructure. The total compute cost was zero, and the analysis is fully reproducible through the Galaxy Europe history sharing system.

Beyond the immediate genomic resource, these assemblies enable marker-assisted selection for the drought and heat tolerance traits that characterize Mahmoudi and the high protein content of Chili. Specific applications include: (1) GWAS panels for Tunisian germplasm, (2) comparative genomics to identify candidate genes for abiotic stress tolerance, and (3) integration into the emerging durum wheat pan-genome.

Several limitations should be noted. First, the sequencing data analyzed here were not generated by the present authors but were obtained from the public deposition of Ayed et al. [2]; raw data provenance and experimental conditions are as described therein. Second, chromosome assignment was performed using sourmash k-mer containment rather than whole-genome alignment tools, which failed due to the 32-bit integer limit of minimap2 (2,147,483,647 bp) on 10+ Gbp polyploid genomes. While sourmash produced confident assignments for A-genome chromosomes, B-genome chromosomes showed elevated repeat co-assignment, most notably Chr3B in Chili (2,731 Mbp vs. 867 Mbp in Svevo v2). These assignments represent best-match chromosome identity; sub-chromosome placement of repetitive scaffolds would require genetic map integration in future work. Third, gene annotation and repeat-element annotation were not performed in the present study and are planned for future work. The chromosome architectures observed here are consistent with our previous population genomic analysis [3], which detected structural variation in Mahmoudi (PCA outlier) and high

Homeologous mapping ambiguity (20–24%) in transcriptome data. The Mahmoudi scaffold with dual Chr2B/Chr3B containment provides the molecular basis for the structural variant signal detected in our short-read analysis, while the Chr3B inflation aligns with the absence of FST selection signals on Chr3B, consistent with a conserved but repeat-rich chromosome. Finally, the slight excess of assembly size over the expected ~10.45 Gbp haploid genome [16] (+3.7% for Chili, +2.4% for Mahmoudi) is interpreted as residual uncollapsed heterozygosity inherent to primary-mode hifiasm assemblies; haplotype-resolved assemblies are also planned as a follow-up.

5. Conclusion

We present two new high-quality, first chromosome-scale-contiguity genome assemblies for the Tunisian durum wheat landraces Chili and Mahmoudi. The assemblies achieve scaffold N50 values approaching individual chromosome sizes, BUSCO completeness scores exceeding 99%, Merqury QV scores of ~68, and validated chromosome assignment to all 14 durum wheat chromosomes, substantially improving upon previously available reference assemblies and providing the first chromosome-scale genomic resources for these North African landraces. By executing the entire workflow on the public Galaxy Europe platform, we demonstrate that reference-quality plant genome assembly is achievable with publicly available data and infrastructure. These resources lay the groundwork for pan-genome construction, marker-assisted breeding, and genome-wide association studies in North African durum wheat germplasm.

Data availability statement

Primary scaffold assemblies are deposited in NCBI GenBank as Third Party Assemblies under BioProject PRJNA1467186 (accessions DBNSEI000000000 for Chili and DBNSEJ000000000 for Mahmoudi). Chromosome-level pseudomolecules are available on Zenodo (DOI: https://doi.org/10.5281/zenodo.20366290). BioSample accessions are SAMN55233146 (Chili) and SAMN55233147 (Mahmoudi). Raw sequencing data are available under NCBI BioProject PRJNA1420514. All assembly and quality assessment analyses were performed on Galaxy Europe (usegalaxy.eu). Chromosome assignment and pseudomolecule construction were performed using locally executed sourmash v4.9.4 and RagTag v2.1.0.

Ethics statement

Not applicable. This study used publicly available sequencing data and did not involve experiments on plants, animals, or humans.

Author contributions

M.G-B.A. and N.E.H.M: Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Writing - Original draft, Writing - Review and editing, Visualization. I.

B.: Investigation, Validation, Writing - Review and editing. All authors read and approved the final manuscript.

Funding

No specific funding was received for this work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank OpenDurumGPT consortium for generating and publicly depositing the raw sequencing data (NCBI BioProject PRJNA1420514) that made this work possible.

Supplementary materials

Supplementary Dataset S1. Complete sourmash chromosome assignment data for both Chili and Mahmoudi pseudomolecules, provided as a single Excel file with two sheets (Chili and Mahmoudi), including scaffold names, sizes, containment indices, and chromosome assignments for all 14 chromosomes (1A–7B) plus unassigned scaffolds (Chr00).

Supplementary Figures S1–S6. Validation figures including chromosome length comparison of Svevo v2 reference, Chili, and Mahmoudi (S1) and collinearity dot plots of Chili and Mahmoudi pseudomolecules against the Svevo v2 reference, A-genome (S2, S4) and B-genome (S3, S5) chromosomes. Collinearity dot plot of the Mahmoudi Chr2B+Chr3B fusion scaffold against concatenated Svevo v2 Chr2B+Chr3B reference.

Submitted as separate files with this manuscript.

References

[1] Trad H, et al. Comparative quality analysis of gluten strength and the relationship with high molecular weight glutenin subunits of 6 Tunisian durum wheat genotypes. Food Sci Biotechnol 2014;23:1363-1370.

[2] Ayed S, Babay E, Ben Amor A, Ben M'Barek S, Djemal R, Ebel C, et al. From local heritage to genomic innovation: whole-genome sequencing of Tunisian durum wheat landraces Chili and Mahmoudi. Zenodo 2026. https://doi.org/10.5281/zenodo.19499032

[3] Gdoura Ben Amor M, Mathlouthi NEH, Belguith I. Unlocking open-access genomic and transcriptomic data: the first bioinformatic exploitation of Tunisian durum wheat landraces Chili and Mahmoudi, pioneering data-driven research in North Africa. bioRxiv 2026. https://doi.org/10.64898/2026.05.08.723814

[4] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 2021;18:170-175. https://doi.org/10.1038/s41592-020-01056-5

[5] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 2011;17:10-12. https://doi.org/10.14806/ej.17.1.200

[6] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013;1303.3997v2.

[7] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-2079. https://doi.org/10.1093/bioinformatics/btp352

[8] Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics 2023;39:btac808. https://doi.org/10.1093/bioinformatics/btac808

[9] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 2013;29:1072-1075. https://doi.org/10.1093/bioinformatics/btt086

[10] Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage. Mol Biol Evol 2021;38:4647-4654. https://doi.org/10.1093/molbev/msab199

[11] Li H. Protein-to-genome alignment with miniprot. Bioinformatics 2023;39:btad014. https://doi.org/10.1093/bioinformatics/btad014

[12] Pierce NT, Irber L, Reiter A, Brooks M, Brown SC, Werneck C, et al. Large-scale sequence comparisons with sourmash. F1000Research 2019;4:1005. https://doi.org/10.12688/f1000research.19675.1

[13] Mazzucotelli E, Llaca V, Fengler K, Charlotte H, Zastrow-Hayes G, Forestan C, et al. Svevo reference genome 2.0: a chromosome-level de novo assembly to support the durum wheat community. Plant Biotechnol J 2025;in press.

[14] Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome biology. 2022 Dec 15;23(1):258.

[15] Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 2020;21:245. https://doi.org/10.1186/s13059-020-02134-9

[16] Maccaferri M, Harris NS, Twardziok SO, et al. Durum wheat genome highlights past domestication signatures and future improvement targets. Nat Genet 2019;51:885-895. https://doi.org/10.1038/s41588-019-0381-3

[17] Chen M, Li J, Zhang Y, Ding Y, Wu M, Li N, et al. Chromosome-level genome assembly of the durum wheat cultivar Langdon. Sci Data 2025;12:376. https://doi.org/10.1038/s41597-025-05724-z