



OPEN

DATA DESCRIPTOR

A Chromosome level assembly of pomegranate (*Punica granatum* L.) variety grown in arid environment

Himanshu V. Patankar^{1,3} , Luis F. Rivera^{1,3} , Fatima Omari Alzahrani², Rod A. Wing¹ & Ikram Blilou¹

The pomegranate (*Punica granatum* L.) is an ancient fruit-bearing tree known for its nutritional and antioxidant properties. They originated from the Middle East in regions having large farms including mountainous regions of Al-Baha in Saudi Arabia. Pomegranates can tolerate arid climates and are considered among the fruits that will play a major role in food security. However, the genomics resources of pomegranate growing in arid regions are scarce. Here, we present a high-quality chromosome-level reference genome using PacBio HiFi long reads. The final assembly was 384.65 Mb with N50 contig size of 43.11 Mb, with 353.42 Mb being anchored on the eight pseudo chromosomes. Annotation revealed that 48.79% of the genome comprises repetitive elements and contains 21,620 protein-coding genes. The new reference genome will contribute to identifying stress resistance traits in pomegranates thriving in arid environments as well as new dietary antioxidants and antimicrobial peptides with pharmaceutical and therapeutic applications.

Background & Summary

Punica granatum is a small deciduous tree widely cultivated for its nutrition-rich fruits¹ that originated in the Middle East and spread across Asia and Europe^{2,3}. The pomegranate fruit has an ancient history and symbolized prosperity and ambition in ancient Egypt⁴. It is a large, round, fibrous berry (Fig. 1A) and it is consumed worldwide, making it one of the most desirable fruits for food industries producing fruit juices⁵. Other than juice, pomegranate fruit waste, especially the nutrient and antioxidant-rich peel, is used in animal feed, baking, the meat industry, and active food packaging⁶. The pharmacological properties, including antioxidants, anti-inflammatory, antidiabetic, antihypertensive, antimicrobial, and anti-tumor of the fruit, make the pomegranate one of the most desirable fruits with respect to healthy eating and medicine^{6,7}.

Pomegranates are mainly cultivated in the Mediterranean countries and the Middle East, where it is considered as a superfood. Worldwide, the pomegranate market is continuously growing, and it is predicted to reach 322.9 million USD by 2026⁸. Saudi Arabia produces about 30,000 tons annually, with most of the production coming from the Al-Baha region of the country⁷. Recently, pomegranates from Saudi Arabia have gained popularity due to their peculiar, sweet taste. A phylogenetic analysis of varieties from Azerbaijan, China, Tunisia, and Saudi Arabia showed that the Saudi variety was closely related to the Chinese pomegranate Dabenzi (Fig. 1I). Recent studies using DNA barcodes for pomegranates grown in Saudi Arabia established an evolutionary relationship between species grown locally and in other regions⁹. Many varieties of pomegranate plants can tolerate a wide range of abiotic stress, including salinity; however, it affects the productivity of the plant^{10,11}. Furthermore, the pomegranate SWEET transporter gene family has been shown to play an important role in both flower development and abiotic stress¹². Despite the stress tolerance of pomegranate in Saudi Arabia, fruit production may drop by about 24% by 2044 due to climate change and soil salinity¹³.

Previous studies on pomegranate focused mainly in conservation, *in vitro* propagation, and the genotyping of Saudi Arabian varieties; however, genetic resources and the genomic data of Saudi pomegranate were unavailable^{9,14}. To fill this void, we provide a high-quality chromosome-level genome assembly of a Saudi pomegranate variety from the Al-Baha region, which is critical to allow studying the genetic diversity and isolating

¹Plant Science Program, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), 23955-6900, Thuwal, Saudi Arabia. ²Department of Biology, Faculty of Sciences, Al-Baha University, Al-Baha, Saudi Arabia. ³These authors contributed equally: Himanshu V. Patankar, Luis F. Rivera. ✉e-mail: himanshu.patanekar@kaust.edu.sa; ikram.blilou@kaust.edu.sa

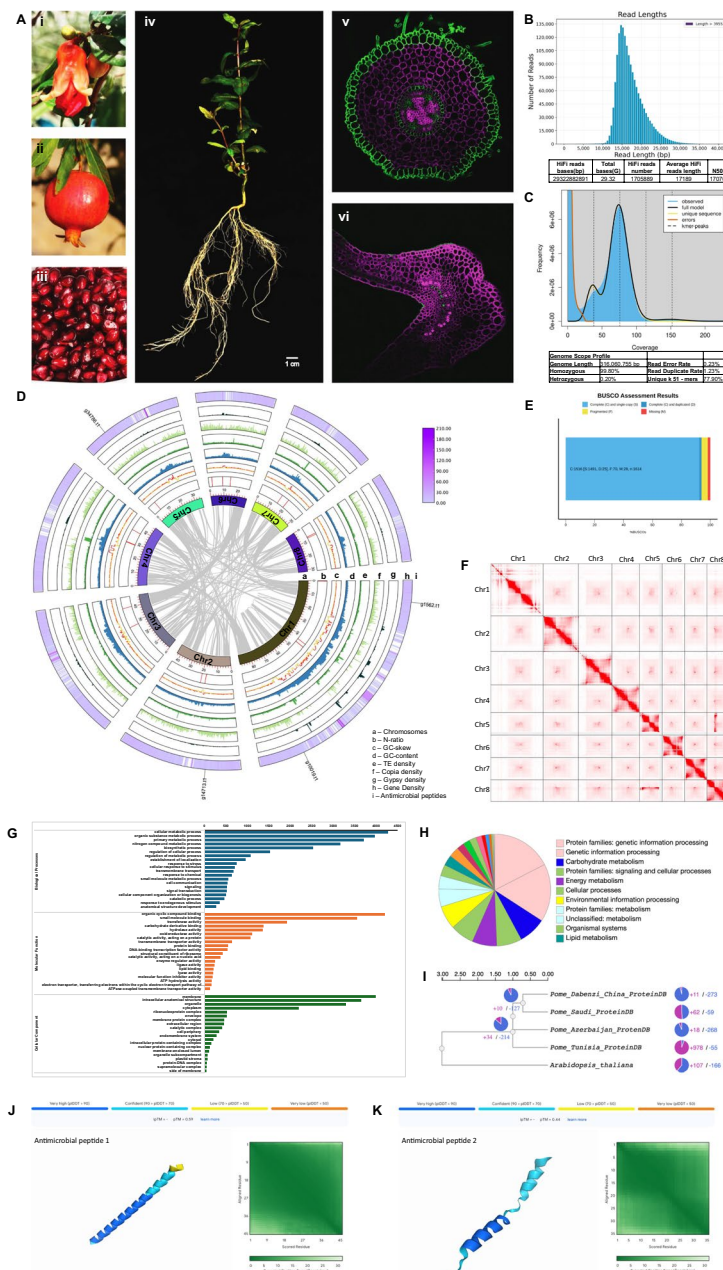


Fig. 1 Images of the Saudi Arabian pomegranate variety sequenced, and statistical and graphical representation of its genomic data. **(A)** KSA pomegranate flower (i), fruit (ii), seeds (iii), seedling (iv), root cross-section (v), and leaf cross-section (vi). **(B)** Quality control data, including read lengths and the statistics of the PacBio HiFi Long reads. **(C)** K-mer (k-51) analysis and the statistics table of genome size estimation by GenomeScope 2.0. The Blue area represents the 51-mer frequency distribution. **(D)** A circular representation of pomegranate genome features, including chromosomes (a), N-ratio (b), GC-skew (c), GC-ratio (d), TE density (e), Ty1/Copia density (f), DIRS1/Gypsy density (g) and gene density (h) antimicrobial peptides (i). **(E)** BUSCO scores of the pomegranate genome assembly. **(F)** Omni-C interaction contact map at the chromosome level. **(G)** Gene Ontology (GO) distribution of the annotated protein-coding genes. **(H)** KEGG annotation distribution across functional categories. **(I)** Phylogenetic analysis of pomegranate varieties from Azerbaijan, China, Tunisia, and Saudi Arabia. **(J,K)** Protein structure prediction by AlphaFold software of high-confidence antimicrobial peptides detected by the Antimicrobial Peptide Database (APD) (J-g34786.t1, K-g10019.t1) (pTM and ipTM scores: the predicted template modeling (pTM) score and the interface predicted template modeling (ipTM) score are both derived from a measure called the template modeling (TM) score. This measures the accuracy of the entire structure. A pTM score above 0.5 means the overall predicted fold for the complex may be similar to the true structure. ipTM measures the accuracy of the predicted relative positions of the subunits within the complex. Values higher than 0.8 represent confident high-quality predictions, while values below 0.6 suggest likely a failed prediction. ipTM values between 0.6 and 0.8 are a gray zone where predictions could be correct or incorrect. TM score is very strict for small structures or short chains, so pTM assigns values less than 0.05 when fewer than 20 tokens are involved; for these cases PAE or pLDDT may be more indicative of prediction quality).

Statistics	Hifiasm assembly	Ragtag Scaffold
Contigs	361	331
Scaffolds	0	30
Chromosomes	0	8
Largest sequence	44,081,199	89,336,501
Total length	384,649,808	384,651,908
N50	23352285	43,112,962
N75	12180252	32,077,463
L50	6	4
L75	12	6
GC (%)	42.14	42.14
N's per 100 kbp	0	0.55

Table 1. Genome assembly statistic based on Hifiasm contig-level and Ragtag scaffold-level assemblies. (the blue curve arrow indicates that the eight chromosome and one chloroplast genome belong to the 30 scaffolds).

stress-tolerant traits in pomegranates. We assembled the pomegranate genome using PacBio HiFi long reads and Omni-C reads. A primary genome assembly was 384.65 Mb with the contig N50 of 43.11 Mb. 353.42 Mb (91.87%) of the assemble anchored onto the eight pseudo-chromosomes. About 50% of the genome was composed of repetitive sequences, of which the long terminal repeats (LTR) and DNA transposons were predominant at 37.29% and 12.31%, respectively. Gene prediction analysis identified 59,019 potential transcripts, of which 21,620 were assigned clear functional roles, including a significant number of protein-coding genes. Importantly, the study also identified four high-confidence antimicrobial peptides that show the potential for applications in the pharmacological and nutraceutical industries. This comprehensive dataset offers a high-quality genome and detailed insights into gene functions and represents a valuable resource for establishing pomegranate research in Saudi Arabia and advancing it internationally.

Methods

Sample and sequencing/sample collection, library construction, and sequencing. Leaves were collected from a healthy voucher pomegranate plant from Al-Baha, Saudi Arabia (19°46'09.1"N 41°27'17.1"E), flash frozen in liquid nitrogen, and stored at -80°C . High molecular weight (HMW) DNA was extracted using Qiagen genomic DNA buffer Set and Qiagen genomic DNA tips 500/G (Qiagen, USA) by the service provided at INRAE-CNRRV, France. HMW DNA quality and quantity were assessed using a FemtoPulse system (Agilent), Qubit fluorometer, and Nanodrop spectrometer. Following HMW DNA extraction, SMRTbell libraries were made and sequenced by PacBio Sequel II at Novogen Co., Ltd., Singapore. Frozen leaf samples were also shipped to Dovetail Genomics, USA for Omni-C library preparation and sequencing.

For RNA sequencing, frozen leaf samples were ground to a fine powder in liquid nitrogen using mortar and pestle, and RNA was extracted using a Qiagen RNeasy Mini kit using manufacturer instructions. IsoSeq libraries were prepared and sequenced at the KAUST Bioscience Core Lab, Saudi Arabia. Briefly, RNA quality and quantity were analyzed using Qubit and Nanodrop; the cDNA was synthesized and amplified using a NEBNext Single Cell/ low Input cDNA Synthesis & Amplification Module. The IsoSeq libraries were prepared using a SMRTbell prep kit 3.0 and sequenced on a PacBio Revio.

Genome survey for genome size estimation. A clean dataset of 55 Gb of PacBio HiFi reads data was used for a quick survey of the pomegranate genome size (Fig. 1B). K-mer ($k=51$) frequency distribution was analyzed using Jellyfish v2.1.4¹⁵, and a histogram was plotted using GenomeScope v2.0¹⁶, where the genome size was estimated to be 316.06 Mb, the heterozygosity rate was 0.2%, and the read error rate was 0.23% (Fig. 1C).

De novo genome assembly. PacBio HiFi reads were used for *de novo* contig level assembly using Hifiasm v0.19.9-r616¹⁷. As a result, the assembled genome size was 384.64 Mb with 361 contigs and a contig N50 size of 23.35 Mb (Table 1). The Hifi contig-level assembly was scaffolded to chromosome-level assembly using Ragtag v2.1.0¹⁸ and a reference pomegranate genome (GCF_007655135.1)¹⁹ as a guide. The ragtag scaffolded assembly size was 384.65 Mb and had an improved contig N50 size of 43.11 Mb (Table 1) (Fig. 1D). The assembled genome's quality was evaluated using QUAST²⁰. The scaffolded assembly was validated with a dot-plot generated by D-GENIES²¹ using Minimap2 v2.28²¹ against the corresponding reference assembly. The completeness of the assembled genome was analyzed by Benchmarking Universal Single-Copy Orthologues (BUSCO) v5.6.1 with the embryophyte_odb10 orthologous database. BUSCO analysis showed that 93.9% of complete Plantae genes were retrieved in the assembled genome (Fig. 1E). Further, the Omni-C reads were aligned to the scaffold-level assembly using Juicer v1.6 and was further manually curated using Juicebox v2.20.00 (Fig. 1F). Subsequently, 91.87% (i.e. 353.42 Mb) of the assembled scaffolds were successfully anchored across eight pseudo-chromosomes (Fig. 1D).

The 30 scaffolds detailed in Table 1 enabled the construction of a chromosome-level genome assembly, utilizing the publicly available genome as a reference guide (GCF_007655135.1). This approach culminated in the resolution of eight chromosomes and a chloroplast genome.

Statistics	#Elements	Length Occupied (bp)	Percent of Genome
LTR elements			
Copia	17,470	13,541,760	3.52%
Gypsy	40,465	46,154,099	12.00%
Unknown	60,739	82916129	21.55%
DNA & MITE			
(DTA, DTC, DTH, DTM, DTT)	59,146	15,328,027	3.99%
Helitrons	122,355	29,742,878	7.73%
Total Repeat elements	299,175	187,683,893	48.79%

Table 2. Repeat annotation statistic of the pomegranate assembly.

Repeat annotation. The Extensive *de-novo* Transposable Element Annotator (EDTA) v2.2.0 software tool was used to detect and annotate transposable elements (TEs) in the assembled genome²² with default setting. Briefly, the EDTA pipeline uses multiple software packaged, i.e. RepeatModeler²³, LTRharvest²⁴, LTR_FINDER²⁵, LTR_retriever²⁶, MITE Tracker²⁷, and HelitronScanner²⁸ for *de novo* TE prediction, and creates a TE library. Finally, RepeatMasker v4.1.5²⁹ was used to identify and annotate the TEs using the EDTA generated TE library. In total, 187.68.19 Mb (48.79%) of the assembled genome consisted of repetitive elements, the majority of which being retroelements (37.07%) and DNA transposons (11.72%) (Table 2) (Fig. 1D).

Gene prediction and functional annotation. Gene prediction was conducted using the AGUSTUS v3.5.0³⁰ software embedded into the Eukaryotic gene finding module of OmicsBox v 3.3.0 (<https://www.biobam.com/omicsbox>), using default parameters in combination with extrinsic IsoSeq evidence³¹ plus *Arabidopsis thaliana* protein sequence data as a reference. AGUSTUS called 59,229 gene sequences, which were then compared with the viridiplantae database using Diamond Blast v2.1.8.162³². BLAST results were loaded into the OmicsBox wrapper for functional annotation and mapping using the InterProScan³³, Gene Ontology (GO) mapping and GO annotation³⁴ modules. Additionally, protein annotation was also performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways³⁵. The InterPro database functionally annotated 31,821 genes, while 22,154 genes were assigned to a specific GO term (Fig. 1G). The KEGG pathway analysis could assign 13377 proteins to their pathway network (Fig. 1H). CAFE 5.0 software, integrated into orthovenn3 (<https://orthovenn3.bioinfotoolkits.net/>), was used for evolutionary phylogenetic analysis³⁶. The phylogenetic tree shows that the Saudi pomegranate variety was closely related to the Debenzi pomegranate variety from China (Fig. 1I).

Additionally, in this study, approximately 460 peptides, each consisting of 50 or fewer amino acid residues, were analyzed using three distinct machine-learning tools publicly available through the CAMP3 platform (<http://www.camp3.bicnirrh.res.in/>)³⁷. Ultimately, four candidate antimicrobial peptides (AMPs) were selected based on the criterion of positively identifying them as AMPs across all predictive tools (Fig. 1J,K). For further validation, these four AMP candidates were manually cross-referenced with the well-known Antimicrobial Peptide Database (APD) (<https://aps.unmc.edu/>), which currently catalogs around 4,000 peptides, over 80% of which are derived from natural sources, as is the case in this study³⁸. Preliminary results revealed that the four candidate peptides exhibited significant homology (greater than 35%) with existing AMPs in the database while showing the potential to represent novel antimicrobial peptides. Additionally, structural predictions were successfully generated using the latest version of AlphaFold (<https://alphafoldserver.com/>), which can be used to test their mode of action. Since many downstream applications, such as molecular dynamics simulations and docking studies, require precise structural information, understanding how these peptides bind to their biological targets (e.g., bacterial cell membranes or intracellular proteins) is critical for elucidating their mode of action. The structural models provided by AlphaFold serve as a foundation for these advanced simulations, facilitating the prediction of binding affinities and other key interactions.

Data Records

Raw data files of PacBio HiFi long reads (SRR30619561)³⁹, Omni-C reads (SRR30619560)⁴⁰, and Iso-Seq reads (SRR30619559)⁴¹ were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA1159017⁴². The genome assembly was deposited in NCBI GenBank accession number JBIMRF000000000⁴³. The genome annotation file is shared on the Figshare database⁴⁴.

Technical Validation

The preliminary analysis to estimate the genome size was performed with clean 55 Gb PacBio HiFi reads K-mer (*k*-51) frequency distribution, analyzed using Jellyfish v2.1.4 and GenomeScope v2.0 (Fig. 1B,C). The genome size was estimated to be 316.06 Mb which was comparable to the current pomegranate NCBI RefSeq assembly (GCF_007655135.1) size of 320.3 Mb. The quality of the assembled genome at the contig level and scaffold level was evaluated using QUAST (Table 1). The completeness of the genome assembly was evaluated by BUSCO v5.6.1 with the embryophyte_odb10 orthologous database. The analysis showed that 93.9% were retrieved from the assembly; the analysis also showed that 92.37% were single copy, complete genes, 1.54% were complete and duplicated genes, 4.33% were fragmented, and 1.73% genes were missing (Fig. 1E). This BUSCO result indicates a high level of completeness and accuracy of the pomegranate genome assembly. Furthermore, the Omni-C interaction contact heatmap indicated a higher interaction intensity along diagonal positions for the eight chromosomes.

Code availability

All software used in this study was implemented using guidelines provided by software manuals; software versions are mentioned in the text. No costume codes were used.

Received: 18 November 2024; Accepted: 18 December 2024;

Published online: 15 January 2025

References

- Melgarejo, P. *et al.* The pomegranate tree in the world: new cultivars and uses. *III International Symposium on Pomegranate and Minor Mediterranean Fruits* 1089, (2013).
- Stone, D. *Pomegranate: A global history*. Reaktion Books (2017).
- Chandra, R. *et al.* Origin, history and domestication of pomegranate. *Fruit, Vegetable and Cereal Science and Biotechnology* 2, 1–6 (2010).
- Heber, D., Schulman, R. N. & Seeram, N. P. *Pomegranates: ancient roots to modern medicine*. CRC press (2006).
- Dhineshkumar, V. & Ramasamy, D. Pomegranate processing and value addition: A review. *Indian Horticulture Journal* 6(1), 1–12 (2016).
- Ain, H. B. U. *et al.* Nutritional importance and industrial uses of pomegranate peel: A critical review. *Food Science & Nutrition* 11(6), 2589–2598 (2023).
- Vučić, V. *et al.* Composition and potential health benefits of pomegranate: a review. *Current pharmaceutical design* 25(16), 1817–1827 (2019).
- Pienaar, L. & Barends-Jones, V. The economic contribution of South Africa's pomegranate industry. *Agriprobe* 18(4), 57–64 (2021).
- Omari Alzahrani, F. *et al.* Employing plant DNA barcodes for pomegranate species identification in Al-Baha Region, Saudi Arabia. *Journal of Umm Al-Qura University for Applied Sciences* 10(1), 136–144 (2024).
- El-Mahdy, M. T., Youssef, M. & Elazab, D. S. In vitro screening for salinity tolerance in pomegranate (*Punica granatum* L.) by morphological and molecular characterization. *Acta Physiologiae Plantarum* 44(2), 27 (2022).
- Liu, C. *et al.* Transcriptomic profiling of pomegranate provides insights into salt tolerance. *Agronomy* 10(1), 44 (2019).
- Kumawat, S. *et al.* Understanding the role of SWEET genes in fruit development and abiotic stress in pomegranate (*Punica granatum* L.). *Molecular Biology Reports* 49(2), 1329–1339 (2022).
- Alghamdi, A. G., Aly, A. A. & Ibrahim, H. M. Effect of climate change on the quality of soil, groundwater, and pomegranate fruit production in Al-Baha Region, Saudi Arabia: a modeling study using SALTMed. *Sustainability* 14(20), 13275 (2022).
- Dessoky, E. *et al.* Genetic studies, in vitro propagation and conservation of some pomegranate cultivars grown in Taif province, Kingdom of Saudi Arabia. *Rev Kasma* 45, 69–86 (2017).
- Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6), 764–770 (2011).
- Vurtture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14), 2202–2204 (2017).
- Cheng, H. *et al.* Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* 18(2), 170–175 (2021).
- Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome biology* 23(1), 258 (2022).
- Luo, X. *et al.* The pomegranate (*Punica granatum* L.) draft genome dissects genetic divergence between soft- and hard-seeded cultivars. *Plant Biotechnology Journal* 18(4), 955–968 (2020).
- Gurevich, A. *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), 1072–1075 (2013).
- Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6, e4958 (2018).
- Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology* 20, 1–18 (2019).
- Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* 117(17), 9451–9457 (2020).
- Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* 9, 1–14 (2008).
- Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research* 35(suppl_2), W265–W268 (2007).
- Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant physiology* 176(2), 1410–1422 (2018).
- Crescente, J. M. *et al.* MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC bioinformatics* 19, 1–10 (2018).
- Xiong, W. *et al.* HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences* 111(28), 10263–10268 (2014).
- Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* 25(1), 4.10. 1–4.10. 14 (2009).
- Hoff, K. J. & Stanke, M. Predicting genes in single genomes with AUGUSTUS. *Current protocols in bioinformatics* 65(1), e57 (2019).
- Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9), 1859–1875 (2005).
- Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature methods* 18(4), 366–368 (2021).
- Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic acids research* 49(D1), D344–D354 (2021).
- Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research* 36(10), 3420–3435 (2008).
- Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology* 428(4), 726–731 (2016).
- Mendes, F. K. *et al.* CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36(22–23), 5516–5518 (2020).
- Waghu, F. H. *et al.* CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic acids research* 44(D1), D1094–D1097 (2016).
- Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic acids research* 44(D1), D1087–D1093 (2016).
- NCBI Sequence Read Archive <http://identifiers.org/insdc.sra:SRR30619561> (2024).
- NCBI Sequence Read Archive <http://identifiers.org/insdc.sra:SRR30619560> (2024).
- NCBI Sequence Read Archive <http://identifiers.org/insdc.sra:SRR30619559> (2024).
- NCBI BioProject <http://identifiers.org/bioproject:PRJNA1159017> (2024).

43. NCBI GenBank <http://identifiers.org/insdc:JBIMRF000000000> (2024).

44. Patankar, H. V. L.F.R., Fatima Omari Alzahrani, Rod A. Wing, Ikram Blilou, A Chromosome level assembly of pomegranate (*Punica granatum* L.) variety grown in arid environment. *Figshare* <https://doi.org/10.6084/m9.figshare.27591690> (2024).

Acknowledgements

The authors thank Dr. Michael Abrouk and Prof. Simon Krattinger for their discussion and valuable suggestions at the initial stage of the work. This study was supported by King Abdullah University of Science and Technology (KAUST) baseline BAS110810101 to Ikram Blilou.

Author contributions

H.P. conceptualized, designed, performed the experiments, analyzed data, and wrote the manuscript. L.R. analyzed data and revised the manuscript. F.O.A. revised the manuscript. R.W. revised the manuscript. I.B. conceptualized, designed the experiment, supervised the work and revised the manuscript

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.V.P. or I.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025