



OPEN ACCESS

EDITED BY

Parthenopi Ralli,
Institute of Plant Breeding and Genetic
Resources, Hellenic Agricultural Organization
(HAO), Greece

REVIEWED BY

Rajesh Kumar Pathak,
Chung-Ang University, Republic of Korea
Prashant Kaushik,
Yokohama Ueki, Japan
Raman Selvakumar,
Indian Agricultural Research Institute (ICAR),
India

*CORRESPONDENCE

Younhee Shin,
✉ yhshin@insilicogen.com

RECEIVED 26 October 2023

ACCEPTED 04 January 2024

PUBLISHED 18 January 2024

CITATION

Park HY, Lim Y-j, Jung M, Sathiyamoorthy S,
Heo SH, Park B and Shin Y (2024), Genome of
Raphanus sativus L. Bakdal, an elite line of large
cultivated Korean radish.
Front. Genet. 15:1328050.
doi: 10.3389/fgene.2024.1328050

COPYRIGHT

© 2024 Park, Lim, Jung, Sathiyamoorthy, Heo,
Park and Shin. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Genome of *Raphanus sativus* L. Bakdal, an elite line of large cultivated Korean radish

Han Yong Park¹, Yu-jin Lim², Myunghee Jung²,
Subramaniam Sathiyamoorthy², Seong Ho Heo^{1,3},
Byeongjun Park^{1,3} and Younhee Shin^{2*}

¹Department of Bioresource Engineering, Sejong University, Seoul, Republic of Korea, ²Research and Development Center, Insilicogen Inc., Yongin-si, Gyeonggi-do, Republic of Korea, ³Institute of Breeding Research, DASAN Co., Ltd., Pyeongtaek, Republic of Korea

KEYWORDS

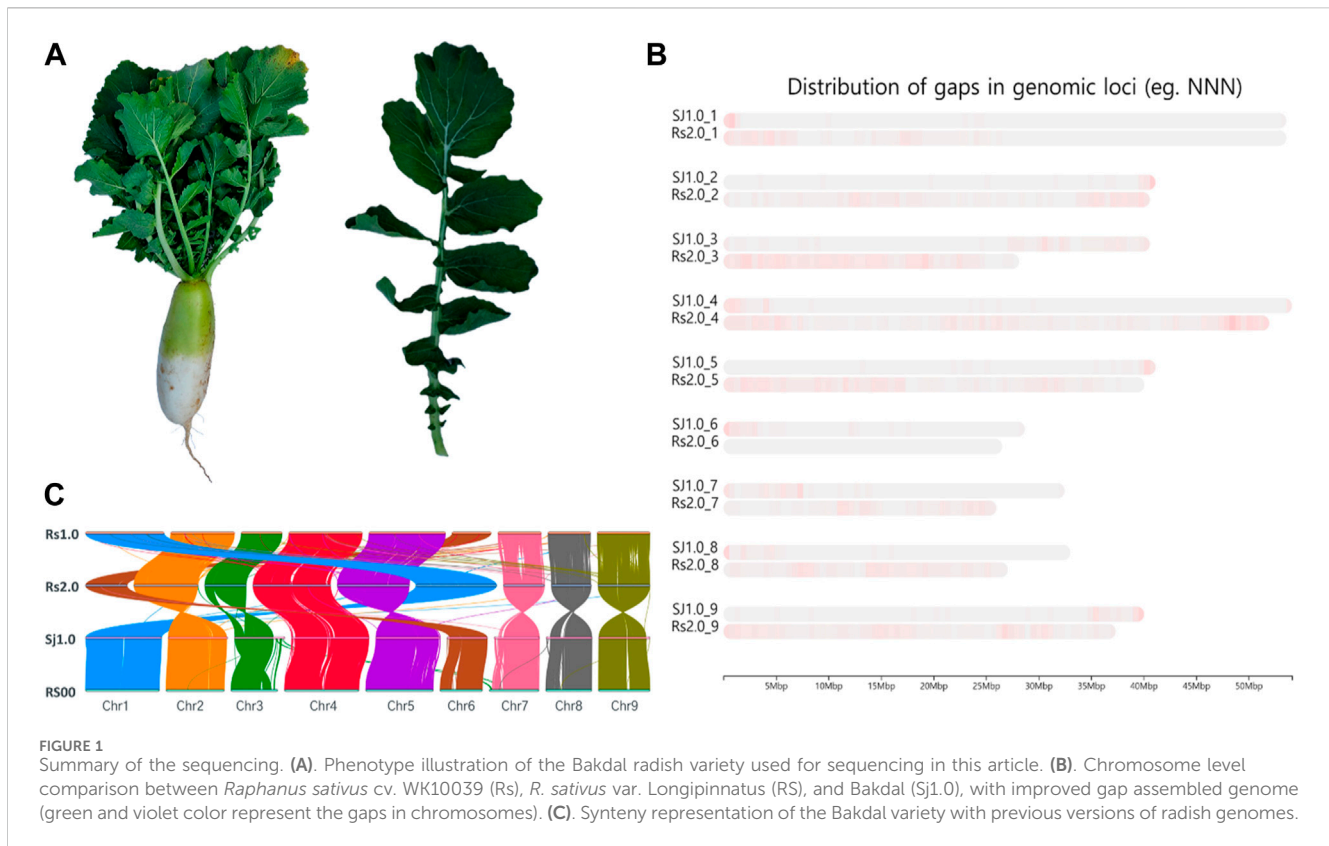
Korean radish, genetics, kimchi, breeding, *Raphanus sativus*

1 Introduction

Kimchi is a common and iconic food in Korean culture and is recognized globally as a healthy food (Surya and Nugroho, 2023). The process of kimchi preparation (kimjang) was included on the Intangible Cultural Heritage of Humanity list by UNESCO in 2013. The *per capita* consumption of kimchi in South Korea surpasses rice consumption (Cha et al., 2023). In Korean culture, fermented food plays a major role and traditional fermentation is one of the major food-processing techniques used to store vegetables during winter.

Kimchi is rich in antioxidants, vitamins, digestive enzymes, and minerals and has anti-cancer, anti-diabetes, and anti-inflammatory properties (Lee et al., 2016). Radish (*Raphanus sativus*) is used to prepare different forms of kimchi, such as *kkakdugi*-kimchi (the second-most consumed kimchi in Korea), *chonggak*-kimchi, and *dongchimi*-kimchi (Song et al., 2021; Surya and Nugroho, 2023). Radishes account for 10% of the total vegetable cultivation land in South Korea (<https://kostat.go.kr>). The burgeoning global popularity of Korean kimchi is increasing the demand for Korean radishes. The need for the sustainable cultivation of radish is growing. Countries like South Korea with small geographic areas are not able to increase the area devoted to agriculture. Instead, crop yield can be increased by establishing new plant traits with the aid of breeding science and technology.

Breeding in the genomic era has become more sophisticated with the development and refinement of technologies that include sequencing. The continuing trend of decreasing cost of genome sequencing and assembly has accelerated the adaptation of digital breeding methods to reveal genetic associations (Marks et al., 2021; Jeon et al., 2023). Vegetable crop varieties have been developed for resistance to various environmental stresses and to meet market demand phenotypes such as taste, aroma, and size. In conventional breeding practices, breeders select desirable traits randomly from nature through manual phenotype assessment. The selected variety is highly documented for a single phenotype, but the abstraction of biological characteristic spectra, such as biotic and abiotic stress resistance, by breeders is not simultaneous with variety/trait selection. Instead, breeders subject the selected varieties to various crossover events to develop new varieties according to the demand and supply in the seed market. This approach is tedious and time-consuming, and can be improved through rapid breeding techniques (Wanga et al., 2021), and genome selection (Budhlakoti et al., 2022). Specifically, breeding through genome selection is comparatively effective, whereas an individual crop has a pan-genome, which includes heterogeneous genomes. The decreased sequencing costs are encouraging plant breeding



scientists to sequence single/target genomes to thousands of genomes aid to construct pan-genomes (Li et al., 2023).

Developing an economically important new radish variety is important in the Korean food industry to retain the aroma, pungency, and taste of kimchi. There are numerous varieties and traits of *R. sativus* and various breeding schemes worldwide aim to develop new traits for this species. The major obstacle in breeding schemes is retaining the crossbreed for multiple generations (i.e., the elite line for each trait/variety) (Kang et al., 2016; Epstein et al., 2023). There are hundreds of economic elite lines available globally; in South Korea, the term “Korean food” specifies that all raw materials used to prepare the food item originate in South Korea (Song et al., 2021). According to the Korean Seed Association, the demand for radish seeds is growing annually. In 2022, the 90.9 tons radish seeds that were produced sold for 445.7 billion Korean won. The chromosomal scale assembled genomes available for the radish crop (i.e., wild species adapted for cultivation that are included in the *Raphanus* pan-genome) are markedly fewer than the existing number of elite lines. In particular, the genome of the Korean radish cultivar *R. sativus* cv. WK10039 was recently decoded using a chromosome scale assembly (Cho et al., 2022). In this study, in addition to the WK10039 genome, we generated the chromosome-scaled genome assembly of *R. sativus* L. Bakdal, a widely cultivated radish line in South Korea which is exclusively used for breeding autumn and summer radishes at commercial breeding companies.

2 Values of the data

The Bakdal genome is an additional genetic resource for elite radish lines. Furthermore, the genome data will provide a basis and

reference for more detailed studies of the genetic variation responsible for stable elite lines. Finally, the genome data may be a valuable resource for conducting comparative analyses of elite and wild varieties of the genus *Raphanus*, which could improve the genome selection process in molecular-assisted breeding.

3 Materials and methods

3.1 Sample collection for genomic DNA extraction

Raphanus sativus L. Bakdal seeds were obtained from the Rural Development Administration, National Institute of Agricultural Science (<https://genebank.rda.go.kr/>). The seeds were maintained under accession number IT100672. Seeds were sown under natural polyhouse conditions established at Sejong University in June 2020. A photograph of the sample is shown in Figure 1A. Four-week-old plant leaves were collected for genome sequencing and immediately stored in liquid nitrogen.

3.2 DNA sequencing and *de novo* genome assembly

Total DNA was isolated from the samples individually according to standard sequencing protocols. DNA was prepared using the TruSeq Nano DNA Prep Kit for Illumina sequencing and the SMRTbell Express Template Preparation Kit (101-357-000; Pacific Bioscience, hereafter PacBio). Each isolated DNA was

TABLE 1 Summary of the sequencing to annotation of the Bakdal chromosomal genome assembly (Sj1.0) along with *Raphanus sativus* cv. WK10039 (Rs2.0).

A. Sequencing and assembly summary		
Descriptions	SJ1.0	Rs2.0
PacBio	57 GB	
Illumina short reads	23 GB	
Illumina Hi-C	30 GB	
Number of Scaffolds	24,155	6,424.00
Total length	398,011,222	433,430,583
N50	40,361,493	37,282,294
Average length	16,477.38	67,534.62
Minimum length	200	698
Maximum length	54,127,229	53,709,999
Number of N	746.468 (0.19%)	269,710 (0.06%)
BUSCO [brassicales_odb10]	C:98.7%[S:90.9%,D:7.8%], F:0.4%, M:0.9%	C:98.6%[S:77.7%,D:20.9%], F:0.3%, M:1.1%
Number of Chromosomes	9	9
Number of Unplaced	24,146	6,413
B. Gene prediction summary		
Number of genes	50,281	56,517
Average gene length (bp)	2,471.04	2,437.05
Gene coverage (%)	31.22	31.78
GC in CDS (%)	46.31	45.99
Exon:Gene	5.32	6.24
Average exon length (bp)	220.16	236.25
Exon coverage (%)	14.8	19.21
Average intron length (bp)	300.58	183.95
Intron coverage (%)	16.41	12.56
Number of rRNAs	966	1,081
Number of tRNAs	1,449	855
C. Functional Annotation Summary		
Database	No. ORFs (%)	
SwissProt	34,514 (68.64)	
KEGG	31,297 (62.24)	
eggong	354 (0.70)	
Pfam	33,663 (66.95)	
GO	32,902 (65.44)	
No hit	11,718 (23.31)	

sequenced using two different sequencing methods: Sequel (PacBio) and Novaseq6000 (Illumina), which are established long- and short-read sequencing techniques. The experiment was performed using DNALink, an authorized service provider in South Korea. The Illumina paired-end and Hi-C sequences were initially subjected to filtering of technical artifacts (i.e., base calling error, Phred quality score [Q] ≤ 20), and adapters using the Trimmomatic-v0.32 method (Bolger et al., 2014). The PacBio reads were error corrected. The

corrected reads were used for the initial *de novo* draft version of the Radish genome with FALCON v1.8.1 haplotype assembler (Chin et al., 2016) and assembly polished with Pilon (Walker et al., 2014). Scaffold artifacts were corrected with 30 GB of Hi-C genome sequence using the HiRise (Putnam et al., 2016) method. Finally, the assembled contigs were scaffolded to the chromosomal scale using the reference radish cultivar, *R. sativus* cv. WK10039 using the RagTag v2.1.0 method (Alonge et al., 2022). The scaffolded contigs

were assessed for completeness using BUSCO v5.4.7, using the Viridiplantae_odb10 reference dataset (Seppey et al., 2019).

3.3 Decontamination and estimation of genome size

Initially, reference databases were prepared in two sequential steps: bacterial contamination and organellar genomes. First, the complete bacterial draft and reference genomes from GenBank were used as references, and the complete organelle genomes (i.e., mitochondria and plastids) were used as references to remove reads from the raw sequence reads. All reference mapping of preprocessed reads was performed using Bowtie2 v2.2.8 (Langmead and Salzberg, 2012). All Illumina preprocessed sequences from the paired-end library were subjected to genome size estimation using the *k*-mer based method previously used for the panda genome (Li et al., 2010). The *k*-mer frequencies (*k*-mer size = 17) were obtained using Jellyfish v2.1.3 (Marçais and Kingsford, 2011). The genome coverage depth was calculated as (*k*-mer Coverage Depth × Average Read Length)/(Average Read Length−*k*-mer size +1). The genome size was calculated as: Total Base Number/Genome Coverage Depth. The *k*-mer Coverage Depth was the major peak in *k*-mer distribution.

3.4 Prediction and classification of repeat regions

Repeat regions in the *R. sativus* L. Bakdal genome were predicted using RepeatModeler (www.repeatmasker.org/RepeatModeler/) and classified into subclasses using the reference Repbase v20.08 database (www.girinst.org/repbase/) (Bao et al., 2015). Finally, the repeats were masked in the genome using RepeatMasker v4.0.5 (www.repeatmasker.org) with RMBlastn v2.2.27+. The results are shown in Supplementary Table S2.

3.5 Gene prediction and annotation

Genes from the *R. sativus* L. Bakdal genome were predicted using an in-house gene prediction pipeline that included an evidence-based gene modeler (EVM), *ab initio* gene modeler, and a consensus gene modeler. The *ab initio* gene modeler and EVM, including exonerate (Slater and Birney, 2005), and AUGUSTUS (Stanke et al., 2006), were trained with several genomes (Figure 1B). The final gene and transcript models were optimized with a consensus gene modeler with EVIDENCEModeler (Haas et al., 2008) and annotated with Trinotate v3.0.1 (Bryant et al., 2017).

3.6 Preliminary analysis

Initially, the size of the *R. sativus* L. Bakdal genome was estimated to be 427.71 (Supplementary Figure S1), based on ~23 GB of short-read Illumina sequences (Table 1A). A 57 GB error corrected long-read sequence was assembled into 398.6 Mb

a 40 MB N50 size gap-free chromosomal scale assembly as illustrated in Supplementary Figure S2 and Supplementary Tables S2, S5. In total, 199 MB (50.03%) assembled genomes were filled with repeat regions, and approximately 30% of the repeats were unclassified into known subclasses of the repeat elements (Supplementary Table S1). The completeness score of the assembled BUSCO genome was 99.9% (Supplementary Figure S3). Furthermore, 50,281 genes were predicted to have an average length of 2,471 bases with other *R. sativus* genomes (Table 1B, Supplementary Figure S4, Supplementary Tables S3–S5). Finally, 68.64% of the genes had homologous sequences in the SwissProt database, and 62.24% of the genes were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. Moreover, 65.44% of the genes had an identified gene ontology (GO). The genic region of the genome was annotated well using reference databases (Table 1C). The Bakdal-assembled genome was mapped to *R. sativus* cv. WK10039 (RS.2.0) (Figure 1B) using the RagTag method to improve scaffold assembly at the chromosome scale. In this study, we generated the genome of the Bakdal line, a commercially important crop widely cultivated in Korea for kimchi production. Compared with RS2.0, the novel genome has additional bases on chromosome three by synteny assessment (Figure 1B). Finally, chromosome one and six were interchanged from previous assemblies (Figure 1C). With the use of this genome research, a full analysis of the genetic information of Bakdal radish has been shown. The most recent discoveries will be of great assistance in the process of incorporating genetic modification into radish cultivars that are predominantly farmed for commercial reasons.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: Sequence Read Archive repository under accession number PRJNA1026765/(10.6084/m9.figshare.24313855).

Author contributions

HP: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing—original draft, Writing—review and editing. Y-JL: Data curation, Formal Analysis, Writing—review and editing. MJ: Data curation, Formal Analysis, Writing—review and editing. SS: Conceptualization, Formal Analysis, Writing—original draft, Writing—review and editing. SH: Methodology, Resources, Validation, Writing—review and editing. BP: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing—review and editing. YS: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Software, Supervision, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work

was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry, and Fisheries (IPET) through the Digital Breeding Transformation Technology Development Program funded by the Ministry of Agriculture, Food, and Rural Affairs (MAFRA) (322063-03-2-SB010).

Conflict of interest

Authors YL, MJ, SS, and YS were employed by Insilicogen Inc. Authors SH and BP were employed by DASAN Co., Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., et al. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23, 258. doi:10.1186/s13059-022-02823-7
- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA.* 2, 6, 11. doi:10.1186/s13100-015-0041-9
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi:10.1016/j.celrep.2016.12.063
- Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., et al. (2022). Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Front. Genet.* 13, 832153. doi:10.3389/fgene.2022.832153
- Cha, J., Kim, Y. B., Park, S.-E., Lee, S. H., Roh, S. W., Son, H.-S., et al. (2023). Does kimchi deserve the status of a probiotic food? *Crit. Rev. Food Sci. Nutr.* 1–14. doi:10.1080/10408398.2023.2170319
- Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods.* 13, 1050–1054. doi:10.1038/nmeth.4035
- Cho, A., Jang, H., Baek, S., Kim, M. J., Yim, B., Huh, S., et al. (2022). An improved *Raphanus sativus* cv. WK10039 genome localizes centromeres, uncovers variation of DNA methylation and resolves arrangement of the ancestral Brassica genome blocks in radish chromosomes. *Theor. Appl. Genet. May.* 135, 1731–1750. doi:10.1007/s00122-022-04066-3
- Epstein, R., Sajai, N., Zelkowski, M., Zhou, A., Robbins, K. R., and Pawlowski, W. P. (2023). Exploring impact of recombination landscapes on breeding outcomes. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2205785119. doi:10.1073/pnas.2205785119
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7. doi:10.1186/gb-2008-9-1-r7
- Jeon, D., Kang, Y., Lee, S., Choi, S., Sung, Y., Lee, T. H., et al. (2023). Digitalizing breeding in plants: a new trend of next-generation breeding based on genomic prediction. *Front. Plant Sci.* 14, 1092584. doi:10.3389/fpls.2023.1092584
- Kang, E. S., Ha, S. M., Ko, H. C., Yu, H.-J., and Chae, W. B. (2016). Reproductive traits and molecular evidence related to the global distribution of cultivated radish (*Raphanus sativus* L.). *Plant Syst. Evol.* 302, 1367–1380. doi:10.1007/s00606-016-1336-0
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9, 357–359. doi:10.1038/nmeth.1923
- Lee, J.-W., Kim, B.-R., Heo, Y., Bae, G.-S., Chang, M. B., and Moon, B. (2016). Feasibility of using kimchi by-products as a source of functional ingredients. *Appl. Biol. Chem.* 59, 799–806. doi:10.1007/s13765-016-0227-y
- Li, N., He, Q., Wang, J., Wang, B., Zhao, J., Huang, S., et al. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* 55, 852–860. doi:10.1038/s41588-023-01340-y
- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., et al. (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* 463, 311–317. doi:10.1038/nature08696
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi:10.1093/bioinformatics/btr011
- Marks, R. A., Hotaling, S., Frandsen, P. B., and VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nat. Plants.* 7, 1571–1578. doi:10.1038/s41477-021-01031-8
- Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., et al. (2016). Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Res.* 26, 342–350. doi:10.1101/gr.193474.115
- Seppy, M., Manni, M., and Zdobnov, E. M. (2019). “BUSCO: assessing genome assembly and annotation completeness,” in *Gene prediction: methods and protocols* (New York, NY: Springer), 227–245. doi:10.1007/978-1-4939-9173-0_14
- Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinforma.* 6, 31. doi:10.1186/1471-2105-6-31
- Song, K., Shin, Y., Jung, M., Subramaniam, S., Lee, K. P., Oh, E. A., et al. (2021). Chromosome-scale genome assemblies of two Korean cucumber inbred lines. *Front. Genet.* 12, 733188. doi:10.3389/fgene.2021.733188
- Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinforma.* 7, 62. doi:10.1186/1471-2105-7-62
- Surya, R., and Nugroho, D. (2023). Kimchi throughout millennia: a narrative review on the early and modern history of kimchi. *J. Ethn. Foods.* 10, 5. doi:10.1186/s42779-023-00171-w
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 9, e112963. doi:10.1371/journal.pone.0112963
- Wanga, M. A., Shimelis, H., Mashilo, J., and Laing, M. D. (2021). Opportunities and challenges of speed breeding: a review. *Plant Breed.* 140, 185–194. doi:10.1111/pbr.12909

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1328050/full#supplementary-material>