台灣農業研究 (J. Taiwan Agric. Res.) 75(2):351–363 (2026)
https://doi.org/10.6156/JTAR.202606_75(2).0013

Research Article

Development of a Genetic Structure-Based Core Collection in *Capsicum* spp. for Efficient Germplasm Utilization

Cheng-Bin Li[1], Jau-Yueh Wang[2], Da-Gin Lin[3], Chin-Chih Chen[4], Ching-Yi Lin[5], Yuan-Kai Tu[6], Cheng-Ping Kuan[6], and Ssu-Yu Lin[1,*]

Abstract

Li, C. B., J. Y. Wang, D. G. Lin, C. C. Chen, C. Y. Lin, Y. K. Tu, C. P. Kuan, and S. Y. Lin. 2026. Development of a genetic structure-based core collection in *Capsicum* spp. for efficient germplasm utilization. J. Taiwan Agric. Res. 75(2):351–363.

We developed a flexible core collection of pepper (*Capsicum* spp.) germplasm from the National Genebank of Taiwan and demonstrated its utility in screening for resistance to pepper mild mottle virus (PMMoV). A total of 1,011 accessions were sequenced, revealing four distinct genetic clusters associated with specific phenotypic traits. Population structure analysis confirmed the strong correspondence between genotypic and phenotypic variation, reflecting the association between population genetic structure and trait expression. Linkage disequilibrium (LD) decay analysis across major *Capsicum* species provided insights into population differentiation and informed single-nucleotide polymorphism (SNP) density requirements for breeding applications. A two-phase core collection strategy was then implemented to balance genetic diversity, representativeness, and experimental feasibility. In the first phase, 250 accessions were selected to enable large-scale studies such as genome-wide association studies (GWAS). This set was further refined in the second phase to 50 accessions, retaining 89.4% of allele coverage. Preliminary application of the core collection to PMMoV resistance screening identified three candidate accessions- Pcc2, Pcc55, and Pcc167- with potential resistance. Overall, this work establishes a genetic structure-informed core collection framework that enhances germplasm utilization, accelerates early-stage resistance evaluation, and provides a foundation for future genomic and breeding research in pepper.

Key words: Pepper, Genetic diversity, Core collection, Pepper mild mottle virus (PMMoV).

INTRODUCTION

Pepper (*Capsicum* spp.) is one of the most important vegetable and spice crops worldwide, valued for both its economic significance and for its nutritional compounds such as capsaicinoids and carotenoids. Breeding has targeted diverse traits- including pungency, fruit morphology,

Received: October 31, 2025; Accepted: January 28, 2026.
* Corresponding author, e-mail: sylin@tari.gov.tw
[1] Assistant Research Fellows, Crop Genetic Resources and Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.
[2] Former Assistant Research Fellow, Crop Genetic Resources and Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.
[3] Contract Associate Research Fellow, Crop Genetic Resources and Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.
[4] Associate Research Fellow, Plant Pathology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.
[5] Associate Research Fellow, Department of Plant Protection, Chiayi Agricultural Experiment Branch, Taiwan Agricultural Research Institute, Chiayi, Taiwan, ROC.
[6] Associate Research Fellows, Crop Genetic Resources and Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.

stress tolerance, and disease resistance- that originate from broad germplasm resources (Nakayama *et al*. 2019; Karim *et al*. 2021). For example, anthracnose resistance in *C*. *annuum* has been introgressed from *C*. *baccatum* and *C*. *chinense*, illustrating the critical role of genetic resources in cultivar development (Barchenger *et al*. 2019; Srivastava & Mangal 2019). Accordingly, national and international genebanks have been established to conserve and provide access to this essential diversity (Pereira-Dias *et al*. 2019; Solomon *et al*. 2019; Tripodi *et al*. 2021).

Although genetic resources provide essential variation for crop improvement, their effective utilization remains constrained by two major challenges: the lack of genetic characterization for many accessions and the difficulty of maintaining large collections. To address these issues, Lee *et al*. (2016) analyzed 3,821 accessions representing 11 species using (SNP) markers and identified 10 genetic clusters and five phylogenetic groups. Similarly, Pereira-Dias *et al*. (2019) evaluated 190 accessions and landraces and classified them into seven clusters. Together, these studies underscore the importance of population structure analysis for the strategic use of germplasm in breeding.

To address the difficulty of maintaining and utilizing excessively large germplasm collections, Brown (1989) proposed the concept of a core collection- a strategically selected subset that captures most of the genetic diversity of the entire collection while minimizing redundancy. By condensing essential variation into a manageable sample size, core collections greatly improve the efficiency of germplasm conservation and utilization. They also provide a practical foundation for genotype-phenotype association studies and trait screening, thereby accelerating breeding progress (Balfourier *et al*. 2007; Richards *et al*. 2009; Mahmoodi *et al*. 2019).

Several core collections of pepper have been established to capture genetic diversity across species and geographic origins. Lee *et al*. (2016) selected 240 accessions from 3,821 using transcriptome-based SNPs, offering a resource for downstream association studies and gene discovery. Similarly, Nicolaï *et al*. (2013) genotyped 1,352 accessions using 28 simple sequence repeat (SSR) markers and constructed a core set of 332 entries that retained 97% of the genetic and phenotypic diversity of *C*. *annuum*. They also proposed nested subsets of varying sizes (8, 16, 32, 64, and 128 accessions), providing flexibility for studies with different resource constraints. This example illustrates that core collections of varying sizes can support experiments at different scales.

Pepper mild mottle virus (PMMoV) is a persistent tobamovirus that poses a serious threat to global pepper production, having been reported in more than 20 pepper-growing countries (Kumari *et al*. 2023; Ochar *et al*. 2023) and causing mottling, leaf deformation, stunting, and yield loss (Elbeshehy *et al*. 2023; Ochar *et al*. 2023). Although resistance is conferred by the *L* gene, the rapid evolution of the virus undermines the durability of these sources, creating an urgent need to identify new resistance (Mongkolporn & Taylor 2011; Barchenger *et al*. 2019; Hur *et al*. 2022). To demonstrate the practical utility of our framework, we analyzed the population structure and linkage disequilibrium of 1,011 pepper accessions, constructed a nested core collection comprising subsets of different sizes, and used PMMoV resistance screening as a case study to evaluate its application. This approach highlights how core collections of varying scales can be applied flexibly to different breeding objectives, while also identifying candidate accessions with potential resistance to PMMoV.

MATERIALS AND METHODS

Plant material and deoxyribonucleic acid (DNA) extraction

A total of 1,054 pepper accessions were used in this study, including 688 *C*. *annuum*, 231 *C*. *baccatum*, 15 *C*. *chinense*, 33 *C*. *frutescens*, 11 *C*. *pubescens*, 21 *C*. *chacoense*, 2 *C*. *eximium*, and one accession each of *C*. *galapagoense*,

C. praetermissum, and *C. tovarii*, along with 50 unclassified accessions. Three plants per accession were grown under standard greenhouse conditions, and young leaves were collected 3 wk after transplanting, flash-frozen in liquid nitrogen, and stored at -80℃. Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), quantified with the Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA), and checked for integrity by 1% agarose gel electrophoresis.

Library preparation and sequencing

Genomic libraries were prepared using double-digest restriction site-associated DNA sequencing (ddRADseq) following Peterson *et al*. (2012) with modifications based on Kao (2016). Genomic DNA was digested with PstI-HF and MspI, ligated to barcoded adapters, PCR-amplified, and size-selected for 250–600 bp fragments using the BluePippin system (Sage Science, Beverly, MA, USA). Library quality was assessed by spectrophotometry, and equal volumes of indexed products (384 accessions per library) were pooled. Paired-end sequencing (150 bp) was performed on an Illumina NovaSeq 6000 platform (S4 flow cell).

Sequence analysis and SNP discovery

Raw sequencing data (.fastq.gz) were processed using "process_radtags" in Stacks v2.61 to demultiplex reads, remove those with incorrect enzyme sites or low quality (Phred < 20), and trim index/adaptor sequences. Clean reads were aligned to the *C. annuum* 'CM334' v1.60 reference genome (Kim *et al*. 2014) using BWA-MEM v0.7.17 (Li 2013). BAM file processing was conducted using SAMtools v1.7 (Li & Durbin 2009), and SNPs were called using the "HaplotypeCaller" module of GATK v4.2.0.0 (McKenna *et al*. 2010). SNP filtering was conducted using VCFtools v0.1.15 with the following criteria: --minDP 3, --minGQ 10, --max-missing 0.5, and --maf 0.05. Samples with an average sequencing depth < 3 (via --depth) were removed, and the remaining 1,011 accessions were used for all subsequent population structure and core collection analyses.

Population structure and genetic diversity analysis

The final SNP dataset (variant call format; VCF) was converted to genomic data structure (GDS) format using the "SNPRelate" package in R. A phylogenetic tree was generated based on the identity-by-state (IBS) matrix using hierarchical clustering and visualized with "ggtree". For population structure analysis, SNP data were converted to STRUCTURE format using PGDSpider v2.1.1.5, and STRUCTURE was run with 10,000 burn-in steps and 100,000 Markov chain Monte Carlo (MCMC) iterations for K = 1–9. The optimal K was determined using the Delta K method with Structure Harvester (Earl & vonHoldt 2012) and visualized in R.

Linkage disequilibrium (LD) decay was analyzed with TASSEL v5 (Bradbury *et al*. 2007), and decay curves were plotted in R. Genotypic principal component analysis (PCA) was performed in "SNPRelate", and the first two principal components were visualized with 90% confidence ellipses. Phenotypic PCA was conducted for 15 standardized traits measured in 250 core accessions; missing data were imputed with "missMDA", and PCA results were visualized in "ggplot2" with trait vectors displayed as biplot arrows.

Core collection construction

Two strategies were used to evaluate core set performance: random sampling and a hybrid A-NE/E-NE optimization using the "corehunter" package in R (De Beukelaer *et al*. 2018). Allele coverage was calculated across various sample sizes (n = 950 to 5) and visualized using LOWESS-smoothing in "ggplot2". Core collection construction was performed in two stages. In Stage 1, 250 accessions were selected from 1,011 using the A-NE/E-NE (0.5/0.5) strategy. Simulations of 1,000 subsets were run to identify accessions with the highest selection frequency. In Stage 2, stepwise nesting was

performed to create smaller core subsets (200, 150, 100, and 50 accessions), all nested within the 250-set, using the same strategy.

PMMoV resistance screening

Two inbred lines, H84 (resistant) and H15 (susceptible), were included as controls. Twenty accessions were randomly selected from the 50-accession core subset, and eight plants per accession were used, with four plants mechanically inoculated and four left uninoculated as healthy controls. At the 4–6 true-leaf stage, plants were mechanically inoculated with PMMoV by gently rubbing carborundum-dusted leaves with sap prepared from infected tissue. Disease severity was assessed 14 d post-inoculation using a four-point disease index (DI): 0 = no symptoms; 1 = mild chlorosis/mottling on new leaves; 2 = chlorosis on new and mature leaves; 3 = severe symptoms on mature leaves. PMMoV infection was confirmed by indirect enzyme-linked immunosorbent assay (ELISA) following Clark & Adams (1977) with modifications (Chen *et al*. 2022), and samples with absorbance values greater than twice those of the healthy controls were considered positive.

RESULTS

Sequencing data analysis and SNP discovery

A total of 1,011 pepper accessions were sequenced using ddRADseq, generating an average of 16.4 million reads per sample with a GC content of 41.2%. After quality filtering, 89.1% of the reads were uniquely mapped to the genome, averaging 13.5 million reads per sample. These reads covered approximately 0.5% of the genome (13 Mb). SNP discovery yielded 235,478 loci, of which 6,437 high-quality SNPs were retained after stringent filtering, corresponding to an average density of one SNP per 543,732 bp. These SNPs spanned all 12 chromosomes of *Capsicum*, providing genome-wide coverage with an average of 536 SNPs per chromosome (Fig. 1).

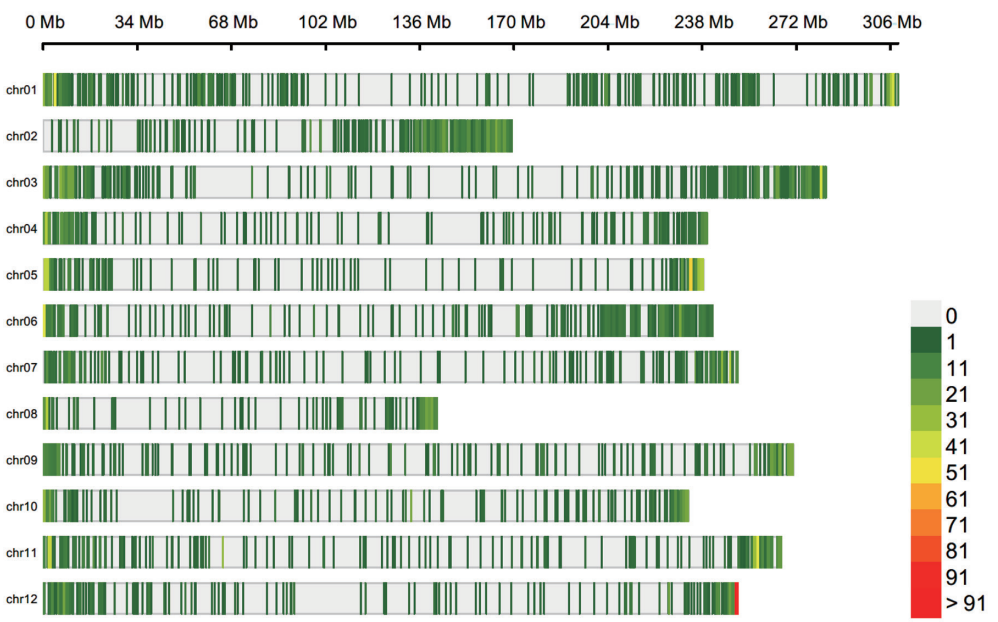

Fig. 1. Genome-wide distribution of single-nucleotide polymorphisms (SNPs) in 1,011 pepper accessions. SNP density across the 12 chromosomes is shown in 1-Mb windows. The color scale represents the number of SNPs per window, with darker green indicating lower density and red indicating higher density.

Population structure analysis

Clustering and phylogenetic analyses revealed clear population structure among the 1,011 accessions. Accessions clustered largely by species, with *C. annuum* forming the largest group, followed by *C. baccatum*, *C. frutescens*, *C. chinense*, and *C. chacoense*. Structure analysis identified four major genetic clusters (ΔK = 4; Fig. 2). Clusters 1 and 3 were composed primarily of *C. annuum*, suggesting the effects of selective breeding, while Cluster 2 included *C. baccatum*, *C. chacoense*, and some *C. pubescens*. Cluster 4 contained mostly *C. frutescens*, *C. chinense*, and a subset of *C. baccatum*. PCA based on genotypes also separated major species and clusters, with *C. annuum* and *C. baccatum* well distinguished, although *C. frutescens* and *C. chinense* showed partial overlap (Fig. 3). These

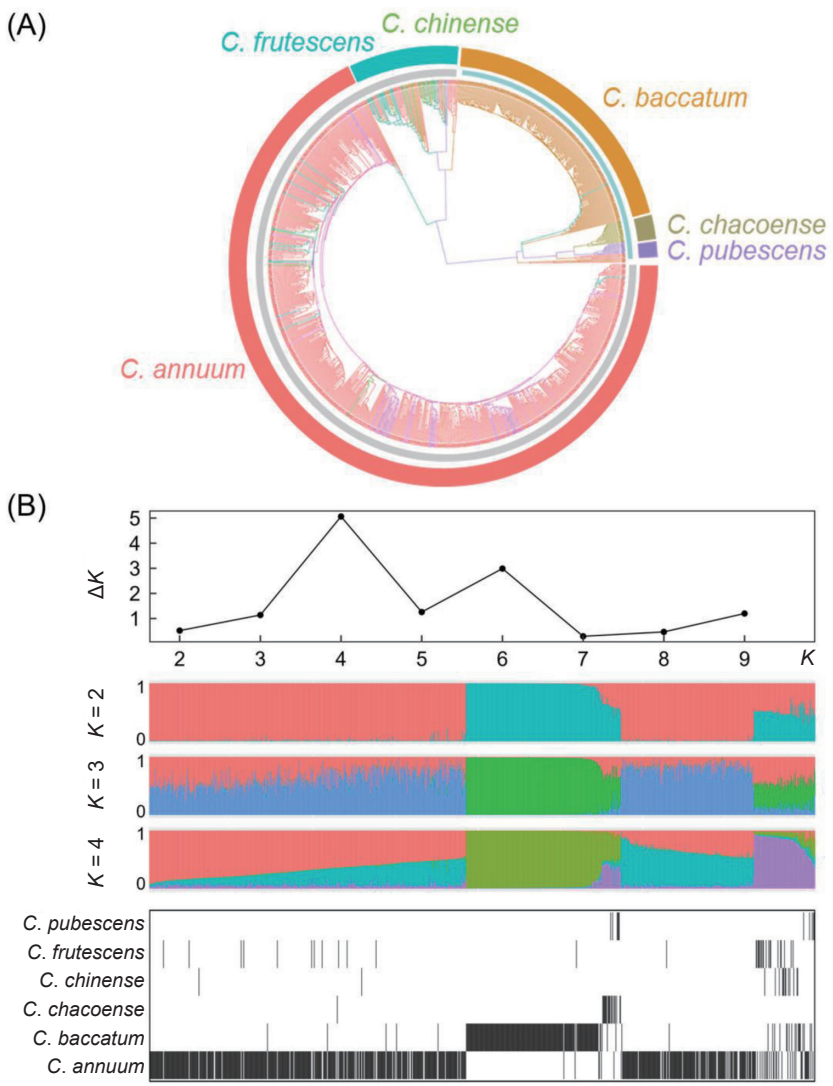

Fig. 2. Population structure of 1,011 pepper accessions. (A) Phylogenetic tree based on the identity-by-state (IBS) matrix, with major *Capsicum* species indicated by colored outer rings. (B) Population structure analysis of K = 2–4 using STRUCTURE. The Delta K plot identifies K = 4 as the optimal number of clusters, and the corresponding bar plots show admixture proportions for each accession.

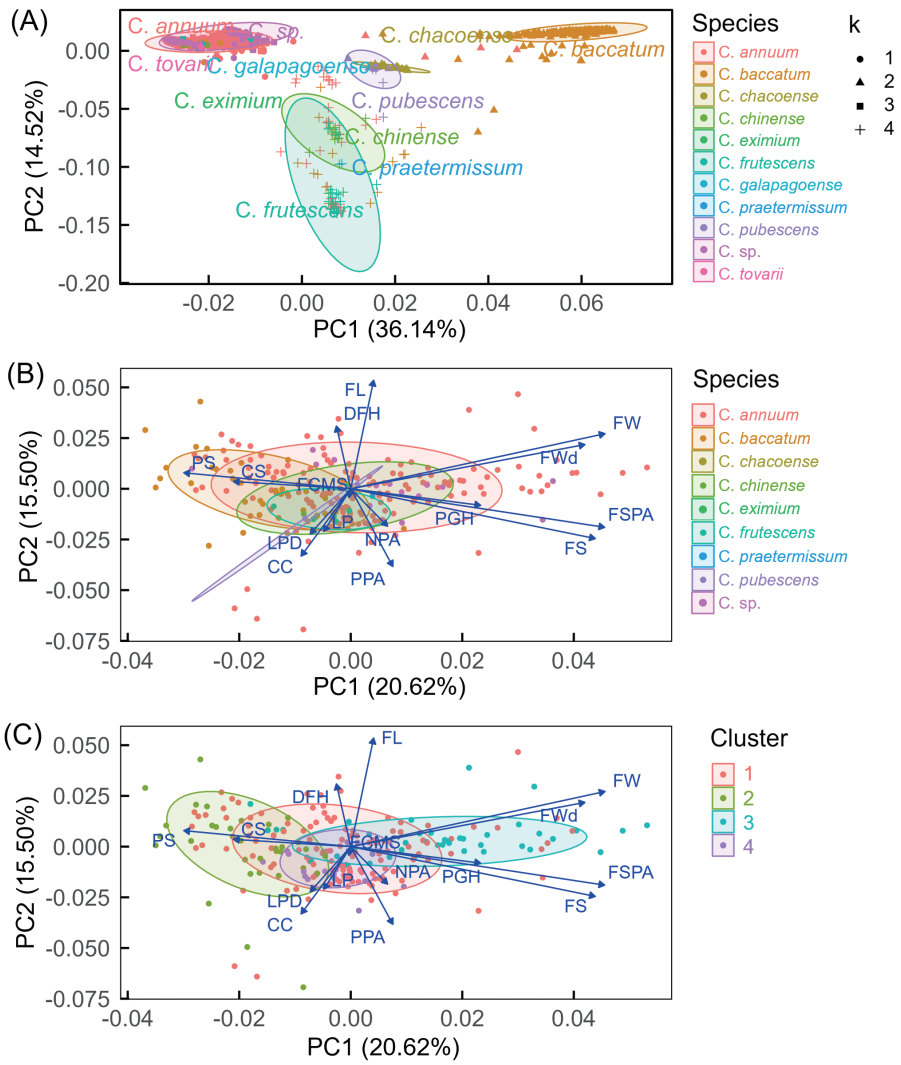

Fig. 3. Principal component analysis (PCA) of the pepper germplasm. (A) Genotypic PCA of 1,011 accessions showing separation of major *Capsicum* species with 90% confidence ellipses. (B) Phenotypic PCA of the 250-accession core set grouped by species. (C) Phenotypic PCA of the 250-accession core set grouped by genetic clusters. Trait loadings are shown as vector arrows, where arrow direction indicates correlation with the principal components and arrow length reflects contribution magnitude. The 15 traits and their abbreviations are: plant growth habit (PGH), plant stature (PS), leaf pubescence density (LPD), leaf pigmentation (LP), pedicel position at anthesis (PPA), number of pedicels per axil (NPA), corolla color (CC), corolla spot (CS), fruit color at mature stage (FCMS), fruit shape (FS), fruit shape at peduncle attachment (FSPA), fruit weight (FW), fruit length (FL), fruit width (FWd), and days from flowering to harvest (DFH).

results are consistent with previous reports (Kim *et al*. 2014).

Linkage disequilibrium decay and population analysis

To evaluate marker density requirements for genome-wide association studies (GWAS), linkage disequilibrium (LD) decay was estimated for the core collection. Across all accessions, LD decayed to an r^2 of 0.2 at 16.71 Mb, indicating that approximately 209 SNPs would be needed to cover the 3.5 Gb pepper genome (Fig. 4).

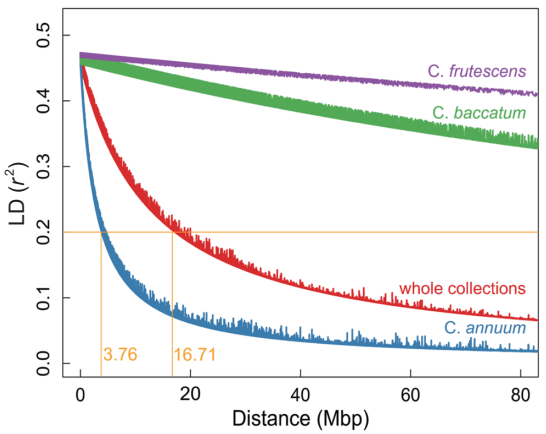

Fig. 4. Linkage disequilibrium (LD) decay in 1,011 pepper accessions. LD decay (r^2) is plotted against physical distance (Mbp) for the full set of accessions and for the three largest species (*Capsicum annuum*, *C. baccatum*, and *C. frutescens*). The horizontal line indicates the r^2 threshold of 0.2, and its intersections with the curves represent the estimated LD decay distances.

Species-specific analyses revealed notable variation: *C. annuum* showed a much shorter LD decay distance of 3.76 Mb, requiring at least 931 SNPs for genome-wide coverage, whereas LD decay thresholds for *C. baccatum* and *C. frutescens* did not reach within the tested range, likely reflecting smaller population sizes and limited recombination sampling. These patterns suggest faster LD decay in *C. annuum*, possibly due to greater recombination activity or stronger historical selection, while the slower decay in *C. baccatum* and *C. frutescens* may result from narrower genetic variation or sampling effects. Overall, the medium-density SNP dataset generated here is sufficient to support GWAS in *Capsicum*, particularly in *C. annuum*.

Construction of core collections with different sample sizes

Building on these population-level insights, we constructed nested core collections of different sizes to address diverse research needs while accommodating experimental limitations such as space and labor. Simulations showed that a 250-accession set captured 92.3% of allelic diversity, while even the smallest 50-accession set retained 89.4% coverage (Fig. 5). Based on these results, a two-phase strategy was implemented: first reducing the full set of 1,011 accessions to 250 (approximately 25%), and then generating nested subsets of 200, 150, 100, and 50 (approximately 5%), with smaller sets nested within larger ones (Table 1). Both phases used a hybrid selection strategy balancing A-NE and E-NE criteria to maximize allelic diversity while minimizing redundancy. Observed genotype coverage closely matched simulation predictions- 92.3% at 250 accessions and 89.4% at 50 accessions- and PCA biplots confirmed that the genetic structure of the subsets mirrored that of the full population (Fig. 6). These results demonstrate that the nested core collections are both representative and scalable, thereby enabling efficient resource allocation for preliminary screening and downstream applications.

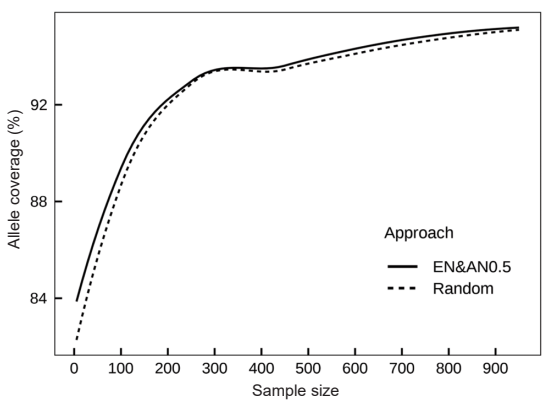

Fig. 5. Allele coverage across different core collection sample sizes. The relationship between sample size and allele coverage (%) is shown for two approaches: CoreHunter3 (solid line) and random sampling (dashed line).

Table 1. Allele coverage of pepper core collections with varying sample sizes.

No. of accessions included in the core collection	Allele coverage
250	92.3%
200	91.8%
150	91.3%
100	90.8%
50	89.4%

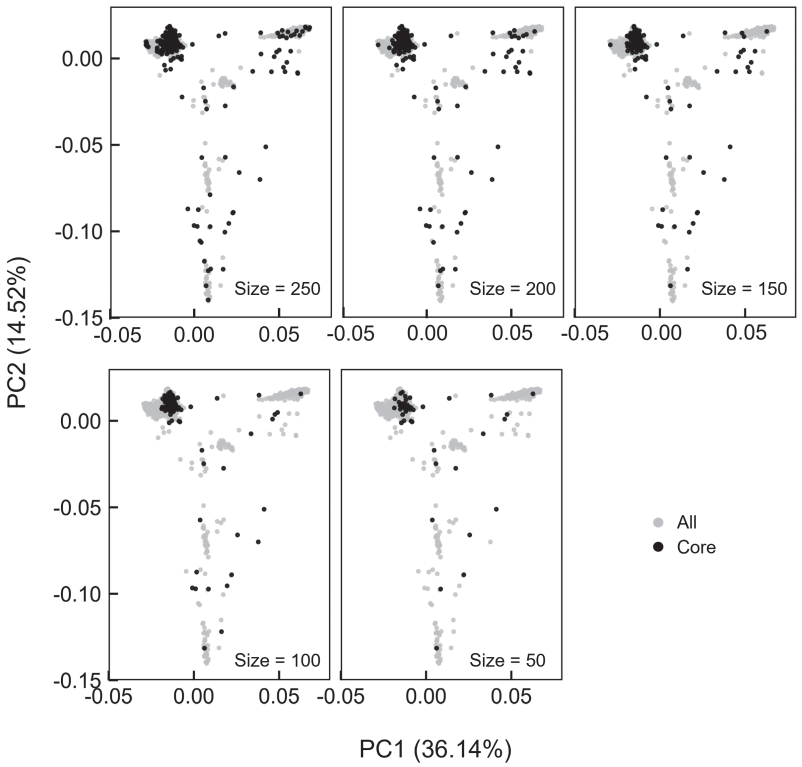

Fig. 6. Genotypic principal component analysis (PCA) of core collections with different sample sizes. PCA biplots of pepper core collections with sample sizes of 250, 200, 150, 100, and 50. Black dots indicate accessions included in each core collection, while gray dots represent the full set of 1,011 accessions.

Phenotypic trait analysis of the pepper core collection

To assess whether the core collection preserved representative phenotypic diversity, we analyzed 15 morphological and agronomic traits in the 250-accession set. In the PCA based on species (Fig. 3B), *C. annuum* displayed the broadest phenotypic variation, reflecting its large population size and domestication history. The *C. baccatum* group was distinguished by strong loadings on plant stature (PS) and corolla spot (CS), while other traits such as fruit weight (FW), fruit width (FWd), fruit shape (FS), and fruit shape at peduncle attachment (FSPA) also contributed to species-level differentiation, although partial overlap remained.

By contrast, PCA based on genetic clusters (Fig. 3C) revealed clearer phenotypic separation. Clusters 1 and 3, both largely composed of *C. annuum*, diverged along FW and FWd, consistent with distinctions between sweet and hot pepper types. Cluster 2, predominantly *C. baccatum*, separated mainly along PS and CS, in agreement with species-level classifications. These results confirm that the core collection captures key phenotypic variation across species and clusters, ensuring the retention of traits relevant to breeding objectives.

Screening of PMMoV-resistant pepper accessions

To demonstrate the practical utility of the constructed core collection in disease resistance screening, PMMoV evaluation was conducted as a case study. Due to the limited capacity of the inoculation platform, which restricted the number of accessions that could be handled simultaneously, 20 *C. annuum*

accessions were randomly selected from the 50-accession core set for evaluation.

The resistant control H84 exhibited low disease index (DI = 0.75 ± 0.96) and ELISA absorbance (0.90 ± 1.06) at 14 d post-inoculation, whereas the susceptible control H15 consistently showed high DI and ELISA values (2.31 ± 0.81 and 2.50 ± 0.58, respectively), confirming the reliability of the assay (Table 2, Fig. 7). Among the tested accessions, Pcc2 (GeneBank ID: 2023A01067) displayed the lowest DI (0.25 ± 0.50) and ELISA value (0.06 ± 0.09), indicating potential resistance to PMMoV. Two additional accessions, Pcc55 (DI = 0.50 ± 0.58; ELISA = 0.70 ± 0.78) and Pcc167 (DI = 1.00 ± 0.82; ELISA = 0.24 ± 0.46), also showed reduced symptom severity and viral accumulation, suggesting their potential as candidate sources of resistance.

Table 2. Resistance analysis of pepper core collection against pepper mild mottle virus (PMMoV).

Name	GeneBank ID	ELISA[z]	DI[y]
H84 (RC[x])	-	0.90 ± 1.06	0.75 ± 0.96
H15 (SC[w])	-	2.31 ± 0.81	2.50 ± 0.58
Pcc2	2023A01067	0.06 ± 0.09	0.25 ± 0.50
Pcc5	2023A01069	0.87 ± 0.70	1.75 ± 1.26
Pcc16	2024A00467	1.29 ± 1.24	1.75 ± 0.96
Pcc53	2023A01139	1.65 ± 1.41	2.50 ± 0.58
Pcc55	2023A01141	0.70 ± 0.78	0.50 ± 0.58
Pcc60	2023A01144	0.25 ± 0.16	1.75 ± 0.96
Pcc63	2023A01147	0.44 ± 0.73	2.00 ± 1.41
Pcc65	2023A01149	0.45 ± 0.73	1.50 ± 1.29
Pcc69	2023A01153	0.77 ± 0.80	1.50 ± 1.29
Pcc72	2023A01156	1.47 ± 1.67	0.75 ± 0.96
Pcc84	2023A01164	1.13 ± 1.36	2.00 ± 0.82
Pcc89	2023A01168	1.74 ± 1.09	1.00 ± 0.82
Pcc91	2023A01169	0.44 ± 0.32	1.50 ± 0.58
Pcc99	2024A00401	1.26 ± 1.30	2.75 ± 0.50
Pcc105	2023A01180	0.88 ± 0.52	2.00 ± 0.82
Pcc146	2023A01213	0.24 ± 0.30	1.50 ± 0.58
Pcc159	2023A01120	1.50 ± 1.01	1.50 ± 1.29
Pcc167	2023A01128	0.24 ± 0.46	1.00 ± 0.82
Pcc168	2023A01129	0.34 ± 0.57	1.25 ± 0.96
Pcc197	2023A01117	1.65 ± 1.25	1.75 ± 0.96

[z] ELISA: enzyme-linked immunosorbent assay.
[y] Disease index (DI) expressed as mean ± standard deviation, measured 14 d post-inoculation.
[x] RC: resistant control.
[w] SC: susceptible control.

DISCUSSION

This study combined genome-wide SNP discovery with population structure and phenotypic analyses to develop nested core collections of pepper. By integrating genomic and phenotypic diversity, the collections provide both a strategic framework for germplasm conservation and practical resources for trait-based breeding.

In this study, a total of 1,011 primary pepper core accessions were genotyped using the ddRADseq, yielding 6,437 high-quality SNP markers distributed across all 12 chromosomes. These markers serve as foundational genomic tools for *Capsicum* research, enabling refined analyses of genetic diversity, species classification, and gene discovery. Population structure analysis revealed clear population differentiation, particularly within *C. annuum*, where gradual transitions between clusters likely reflect historical selection for sweet and chili types. Integration of clustering results with phenotypic data further supported this distinction, highlighting the utility of genetic structure information in guiding parent selection for trait-oriented breeding.

In contrast, species such as *C. baccatum* and *C. frutescens* exhibited more homogeneous genetic profiles, suggesting reproductive isolation or less intensive selection. Nevertheless, several admixed accessions were identified, representing promising materials for interspecific introgression, especially for the transfer of disease resistance traits.

LD decay patterns also have important implications for breeding strategies. The rapid decay observed in *C. annuum* indicates its suitability for high-resolution GWAS, but emphasizes the need for dense marker coverage and precise phenotyping. By contrast, the slower decay in *C. baccatum* and *C. frutescens* suggests that lower marker densities may suffice for trait mapping in these species, which

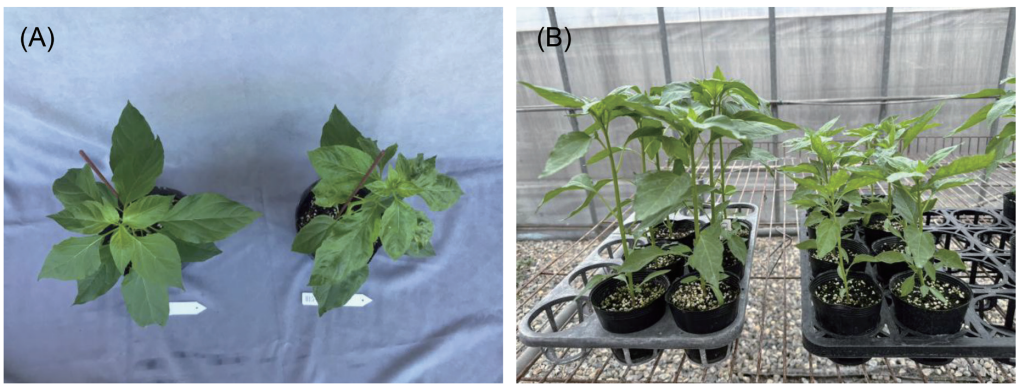

Fig. 7. Symptoms caused by pepper mild mottle virus (PMMoV) on the susceptible control H15 at 14 d post-inoculation. (A) Mosaic symptoms observed in inoculated plants (right) compared with non-inoculated controls (left). (B) Stunting was observed in inoculated plants (right) compared with non-inoculated controls (left).

could reduce genotyping costs. Together, these insights help optimize marker-assisted strategies across different pepper gene pools, particularly when introgressing traits from wild or minor species.

To address practical limitations in germplasm evaluation, we developed a nested core collection consisting of subsets of 250, 200, 150, 100, and 50 accessions. This stepwise design balances allele diversity, representativeness, and scalability, ensuring that researchers can select a core size appropriate to their experimental capacity. Even the smallest subsets preserved most of the genetic and phenotypic variation, demonstrating that compact cores can still serve as effective proxies for the full population. Importantly, phenotypic analyses confirmed that key agronomic distinctions, such as those between hot and sweet *C. annuum*, were retained, underscoring the value of the cores for both germplasm conservation and trait-based selection.

To illustrate the utility of the nested core collection under practical breeding constraints, the PMMoV assay in this study was intentionally designed as an early-stage, exploratory screening rather than as a definitive resistance validation. Accordingly, result interpretation focused on overall trends and relative performance among accessions, rather than on strict hypothesis testing.

Although differences between the resistant (H84) and susceptible (H15) controls did not reach statistical significance, the observed mean differences in both DI, reflecting external symptom expression, and ELISA-based measurements of viral accumulation were directionally consistent with their established resistance phenotypes. The lack of statistical significance is most plausibly attributable to limited replication (four inoculated plants per line), together with variability inherent to mechanical inoculation and the distinct measurement characteristics of DI and ELISA.

DI and ELISA capture complementary but non-equivalent aspects of host-virus interactions. DI represents symptom severity on an ordinal scale with constrained resolution and potential ceiling effects, whereas ELISA quantifies viral accumulation at the tissue level and is inherently more sensitive to quantitative variation. Consequently, greater variability in ELISA relative to DI is expected in early-stage screening assays, particularly for accessions exhibiting partial resistance or tolerance, where symptom expression does not necessarily scale linearly with viral load.

Within this framework, accessions were compared with the resistant and susceptible controls, focusing on those with relatively low DI scores and reduced viral accumulation.

Under these conditions, Pcc2, Pcc55, and Pcc167 were among the accessions exhibiting this response pattern. However, these results should be interpreted with caution, and further evaluation with increased replication and repeated inoculation trials is required before any conclusion can be made regarding their resistance.

Overall, this study presents a strategic framework for constructing and applying nested pepper core collections based on genetic structure and phenotypic diversity. The established collections enhance the efficiency of trait screening and breeding while contributing to the long-term conservation and utilization of genetic resources. The developed pepper core collection can be applied to support marker-assisted selection, GWAS, and the development of elite cultivars under practical breeding constraints.

CONCLUSIONS

This study developed a nested core collection of pepper based on genome-wide SNPs and population structure, providing a scalable resource that balances genetic diversity, representativeness, and experimental feasibility. Validation with phenotypic traits and a case study on PMMoV resistance confirmed its potential for efficient trait screening. The developed core sets offer a practical tool for conserving genetic diversity and can be readily applied to support breeding programs, marker-assisted selection, and genome-wide studies under practical constraints.

ACKNOWLEDGMENTS

The authors would like to thank the Ministry of Agriculture for funding support through projects 110AS-4.6.1-CI-C9 and 111AS-4.6.1-CI-C7. The pepper germplasm used in this study was kindly provided by the National Plant Genetic Resources Center of the Taiwan Agricultural Research Institute and the Genebank of World Vegetable Center. We appreciate their generous contributions, which were essential for the success of this research.

REFERENCES

Balfourier, F., V. Roussel, P. Strelchenko, F. Exbrayat-Vinson, P. Sourdille, G. Boutet, ... G. Charmet. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor. Appl. Genet. 114:1265–1275. doi:10.1007/s00122-007-0517-1

Barchenger, D. W., P. Naresh, and S. Kumar. 2019. Genetic resources of *Capsicum*. p.9–23. *in*: The *Capsicum* Genome. (Ramchiary, N. and C. Kole, eds.) Springer. Cham, Switzerland. 232 pp. doi:10.1007/978-3-319-97217-6_2

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Brown, A. H. D. 1989. Core collections: A practical approach to genetic resources management. Genome 31:818–824. doi:10.1139/g89-144

Chen, C. C., F. L. Chiang, Y. H. Cheng, and J. Y. Liao. 2022. Improvement of wide-spectrum detection of different potyviruses infecting passion fruit by the mixture of two kinds of polyclonal antibodies against east Asian passiflora virus. J. Taiwan Agric. Res. 71:101–121. (in Chinese with English abstract) doi:10.6156/JTAR.202206_71(2).0002

Clark, M. F. and A. N. Adams. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 34:475–483. doi:10.1099/0022-1317-34-3-475

De Beukelaer, H., G. F. Davenport, and V. Fack. 2018. Core Hunter 3: Flexible core subset selection. BMC Bioinformatics 19:203. doi:10.1186/s12859-018-2209-z

Earl, D. A. and B. M. vonHoldt. 2012. Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361. doi:10.1007/s12686-011-9548-7

Elbeshehy, E. K. F., W. M. Hassan, and A. A. Baeshen. 2023. Controlling pepper mild mottle virus (PMMoV) infection in pepper seedlings by use of chemically synthetic silver nanoparticles. Molecules 28:139. doi:10.3390/molecules28010139

Hur, O. S., H. R. Kwak, N. Y. Ro, Y. Choi, S. Lee, A. Hwang, ... B. S. Hahn. 2022. Resistance screening of *Capsicum* germplasm to pepper mild mottle virus (PMMoV) pathotypes $P_{1,2}$ and $P_{1,2,3}$. Korean J. Breed. Sci. 54:1–7. (in Korean with English abstract) doi:10.9787/KJBS.2022.54.1.1

Kao, C. 2016. Construction of a high density linkage map

with double digest RADseq and mapping of restorer-of-fertility gene in pepper (*Capsicum annuum* L.). Master Thesis. Department of Agronomy, National Taiwan University. Taipei, Taiwan. 94 pp. (in Chinese with English abstract) doi:10.6342/NTU201610013

Karim, K. M. R., M. Y. Rafii, A. B. Misran, M. F. B. Ismail, A. R. Harun, M. M. H. Khan, and M. F. N. Chowdhury. 2021. Current and prospective strategies in the varietal improvement of chilli (*Capsicum annuum* L.) specially heterosis breeding. Agronomy 11:2217. doi:10.3390/agronomy11112217

Kim, S., M. Park, S. I. Yeom, Y. M. Kim, J. M. Lee, H. A. Lee, … D. Choi. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat. Genet. 46:270–278. doi:10.1038/ng.2877

Kumari, N., V. Sharma, P. Patel, and P. N. Sharma. 2023. Pepper mild mottle virus: A formidable foe of *Capsicum* production- A review. Front. Virol. 3:1208853. doi:10.3389/fviro.2023.1208853

Lee, H. Y., N. Y. Ro, H. J. Jeong, J. K. Kwon, J. Jo, Y. Ha, … B. C. Kang. 2016. Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. BMC Genet. 17:142. doi:10.1186/s12863-016-0452-8

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. doi:10.48550/arXiv.1303.3997

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324

Mahmoodi, R., M. R. Dadpour, D. Hassani, M. Zeinalabedini, E. Vendramin, S. Micali, and F. Z. Nahandi. 2019. Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. Sci. Hortic. 249:439–448. doi:10.1016/j.scienta.2019.02.017

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, … M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110

Mongkolporn, O. and P. W. J. Taylor. 2011. *Capsicum*. p.43–57. *in*: Wild Crop Relatives: Genomic and Breeding Resources. (Kole, C., ed.) Springer. Heidelberg, Germany. 282 pp. doi:10.1007/978-3-642-20450-0_4

Nakayama, T., S. Takahashi, and T. Waki. 2019. Formation of flavonoid metabolons: Functional significance of protein-protein interactions and impact on flavonoid chemodiversity. Front. Plant Sci. 10:821. doi:10.3389/fpls.2019.00821

Nicolaï, M., M. Cantet, V. Lefebvre, A. M. Sage-Palloix, and A. Palloix. 2013. Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. Genet. Resour. Crop Evol. 60:2375–2390. doi:10.1007/s10722-013-0006-0

Ochar, K., H. C. Ko, H. J. Woo, B. S. Hahn, and O. Hur. 2023. Pepper mild mottle virus: An infectious pathogen in pepper production and a potential indicator of domestic water quality. Viruses 15:282. doi:10.3390/v15020282

Pereira-Dias, L., S. Vilanova, A. Fita, J. Prohens, and A. Rodríguez-Burruezo. 2019. Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). Hortic. Res. 6:54. doi:10.1038/s41438-019-0132-8

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135. doi:10.1371/journal.pone.0037135

Richards, C. M., G. M. Volk, P. A. Reeves, A. A. Reilley, A. D. Henk, P. L. Forsline, and H. S. Aldwinckle. 2009. Selection of stratified core sets representing wild apple (*Malus sieversii*). J. Amer. Soc. Hortic. Sci. 134:228–235. doi:10.21273/JASHS.134.2.228

Solomon, A. M., K. Han, J. H. Lee, H. Y. Lee, S. Jang, and B. C. Kang. 2019. Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. PLoS One 14:e0216886. doi:10.1371/journal.pone.0216886

Srivastava, A. and M. Mangal. 2019. Capsicum breeding: History and development. p.25–55. *in*: The Capsicum Genome. (Ramchiary, N. and C. Kole, eds.) Springer. Cham, Switzerland. 232 pp. doi:10.1007/978-3-319-97217-6_3

Tripodi, P., M. T. Rabanus-Wallace, L. Barchi, S. Kale, S. Esposito, A. Acquadro, … N. Stein. 2021. Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. Proc. Natl. Acad. Sci. U.S.A. 118:e2104315118. doi:10.1073/pnas.2104315118

基於遺傳結構之番椒核心種原建立與應用

李承彬[1] 王昭月[2] 林大鈞[3] 陳金枝[4] 林靜宜[5] 杜元凱[6] 關政平[6] 林思妤[1,*]

摘要

李承彬、王昭月、林大鈞、陳金枝、林靜宜、杜元凱、關政平、林思妤。2026。基於遺傳結構之番椒核心種原建立與應用。台灣農業研究 75(2):351–363。

本研究旨在建立以遺傳結構為基礎之番椒 (*Capsicum* spp.) 核心種原，以促進遺傳資源之保存與應用。本研究自國家種原庫蒐集 1,011 份番椒材料，採用雙重酶切 DNA 定序技術進行基因體分析，共獲得 6,437 個高品質單核苷酸多型性位點。族群結構分析顯示番椒種原可分為四大遺傳族群，其基因型與外表型變異具高度對應關係，反映族群遺傳結構與性狀表現之關聯性。進一步之連鎖不平衡分析顯示番椒種間之基因體分化特徵，可作為後續全基因體關聯性研究之分子標誌密度參考。基於此些遺傳背景資訊，本研究採雙階段策略建立番椒核心種原，首先建立涵蓋 250 個種原之番椒核心種原，再逐步縮減為涵蓋 200、150、100 及 50 個種原之番椒核心種原，此些核心種原保留 89.4% 以上之等位基因多樣性。為驗證核心種原之實用性，本研究將番椒核心種原應用於番椒微斑病毒 (pepper mild mottle virus; PMMoV) 抗性評估，結果顯示 Pcc2、Pcc55 與 Pcc167 等種原具抗病潛力。本研究建立之番椒核心種原為番椒遺傳資源之保存與應用奠定新的基礎，亦可作為後續抗逆境篩選與全基因體關聯分析等研究之重要材料。

關鍵詞：番椒、基因多樣性、核心種原、番椒微斑病毒。

投稿日期：2025 年 10 月 31 日；接受日期：2026 年 1 月 28 日。

* 通訊作者：sylin@tari.gov.tw
[1] 農業部農業試驗所遺傳資源與生物技術組助理研究員。臺灣 臺中市。
[2] 農業部農業試驗所遺傳資源與生物技術組前助理研究員。臺灣 臺中市。
[3] 農業部農業試驗所遺傳資源與生物技術組聘用副研究員。臺灣 臺中市。
[4] 農業部農業試驗所植物病理組副研究員。臺灣 臺中市。
[5] 農業部農業試驗所嘉義農業試驗分所植物保護系副研究員。臺灣 嘉義市。
[6] 農業部農業試驗所遺傳資源與生物技術組副研究員。臺灣 臺中市。