Physiology and Molecular Biology of Plants (March 2025) 31(3):477–491
https://doi.org/10.1007/s12298-025-01576-4

RESEARCH ARTICLE

Metabolite profiling reveals differential accumulation of secondary metabolites related to flavour and colour across four heirloom chilli landraces

Aheibam Loyanganba Meitei[1] · Kalenahalli Yogendra[2] · Hemalatha Sanivarapu[2] · Ngasepam Tombisana Meetei[1] · Mayank Rai[1,3] · Wricha Tyagi[1,2]

Received: 12 December 2024 / Revised: 17 March 2025 / Accepted: 18 March 2025 / Published online: 6 April 2025

Abstract

Chillies from Northeast India exhibit wide variability in fruit morphology, pungency, bearing habit and crop duration. An untargeted metabolite profiling using LC-HRMS of four 'heirloom' pungent landraces viz. Naga chilli (AL-1), Dalle khursani (AL-2), Sohmynken khnai (AL-3), and J-41(B) was performed and compared with Kashi anmol (KA). While AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is categorised as *C. chinense* and AL-3 is *C. frutescens*. A total of 6990 consistent peaks of monoisotopic masses were detected, out of which 2702 metabolites were identified using accurate mass error < 10 ppm. A higher number of differentially accumulated metabolites were seen in J-41(B) versus AL-3 (1376), followed by J-41(B) versus AL-2 (1365), J-41(B) versus AL-1 (1257), KA versus AL-2 (649), AL-3 versus KA (616), KA versus AL-1 (594) and J-41(B) versus KA (413). Variation among species was higher than variation within species. Pathway analysis identified fatty acid, carotenoid, flavonoid and capsaicinoid as key pathways. We identified eight major categories of metabolites, including fatty acids, sterol lipids, and flavonoids, which together account for over 70% of the significantly expressed metabolites across the genotypes. This study explores untargeted metabolites in various chilli species, offering insights into the biochemical and molecular mechanisms which may play a role in governing important fruit traits. Identification of key metabolites and underlying alleles for twenty-one genes across three pathways (flavonoid, capsaicinoid and carotenoid) suggests that the metabolites and associated alleles identified in this study can be used as biomarkers for further characterization of these heirloom chilli and could provide distinct parameter(s) in distinguishing improved cultivars from landraces. This will contribute towards breeding programs in aiding selection of fruits of the desirable traits.

Keywords *Capsicum* · Metabolites · Carotenoid · Flavonoid · Allele mining

Introduction

Chillies (*Capsicum* spp.) or peppers are one of the most important spice crops belonging to the Solanaceae family. They have a complex genome and are known for their diversity of fruit shape, size, colour and pungency. Chillies are used as vegetables, seasoning, ornamental plants, and medicinal crops. They are good sources of various beneficial compounds, such as ascorbic acid (vitamin C), carotenoids (provitamin A), tocopherols (vitamin E), flavonoids, and capsaicinoids. Several health benefits, such as anti-inflammatory, antidiabetic, antimicrobial, anticholesteremic, anticlotting and antioxidant activities (Wahyuni et al. 2013a, b; Chamikara et al. 2016) are also reported. There are more than 30 species of *Capsicum*, and the domesticated and economically important species are *Capsicum annuum,*

✉ Wricha Tyagi
wricha.tyagi@icrisat.org

[1] School of Crop Improvement, College of Post Graduate Studies in Agricultural Sciences (CPGSAS), Central Agricultural University (Imphal), Umiam, Meghalaya 793103, India

[2] Research Program- Accelerated Crop Improvement (ACI), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana 502324, India

[3] Post Graduate College of Agriculture, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Pusa, Samastipur, Bihar 848125, India

C. baccatum, C. chinense, C. frutescens and *C. pubescens* (Perry et al. 2007; Lee et al. 2020). India in 2021–2022 had an annual chilli production of 18.74 lakh tonnes, of which north-eastern states contributed only 2.73% (Apeda et al. 2021), despite the region harbouring the most pungent chilli in the country, with diverse shapes and sizes.

Significant diversity in fruit shape, spiciness, growth patterns, and duration has been reported for chillies cultivated in Northeast India (Yumnam et al. 2012). Genotypes belonging to different species with characteristic morphological and growing habits also differ in their primary and secondary metabolites (Wahyuni et al. 2011, 2013a, b). Different metabolites (volatile and non-volatile compounds) play important roles in our dietary choice of food flavours and taste, ranging from sweetness to spiciness. Metabolites in chilli have been characterized by targeting specific compounds using either gas chromatography (GC) (Gaur et al. 2016) or high-performance liquid chromatography (HPLC) analysis (Wahyuni et al. 2011; Naseem et al. 2024). Metabolite profiling for 136 chilli genotypes using GC–MS, which can only identify volatile compounds, revealed 56 metabolites only (Gaur et al. 2016). This suggests a need for a better molecular understanding of traits of commercial relevance using an untargeted approach.

Among the landraces, *C. chinense* (King chilli/Bhoot Jolokia/Naga chilli/Umorok), *C. annuum* (Dalle khursani; J41(B)) and *C. frutescens* (Sohmynken khnai) have higher pungency and capsaicin content in comparison with *C. annuum* (Sanatombi and Sharma 2008; Islam et al. 2015). In addition to this, *C. annuum* (Dalle khursani) is an allopolyploid having 2n = 4x = 48 with higher pungency than their genetically related cultivars of *C. annuum* (Jha et al. 2017; Colney et al. 2018). Fruit characters and pungency are the main basis on which chilli is classified into different commercial values. Thus, distinguishing parameters are essential to avoid the grouping of improved cultivars and landraces into one group based on mutual resemblance (Geleta et al. 2005). Therefore, to get a better insight of these complex but crucial fruit-related traits, the untargeted metabolite profiles of mature ripe fruits of four distinct 'heirloom' landraces of the Northeastern region of India (King chilli, Dalle khursani, J-41(B) and Sohmynken khnai) (Yumnam et al. 2012; Colney et al. 2018) were analysed and compared with Kashi Anmol (commercial cultivar). These landraces were selected based on variation in geographical origin, fruit morphology (size, shape, texture and colour) and distinct flavour. Earlier studies suggested that it is difficult to demarcate the individual species of *C. annuum*, *C. frutescens* and *C. chinense* based on the morphological descriptors (Pickersgill et al. 1979). The diversity assessment with molecular markers revealed that *C. annuum* is most closely related to *C. frutescens*, followed by *C. chinense* and more distantly related to *C. pubescens* (Jeong et al. 2010). We have previously reported that out of the total variation detected across 56 diverse genotypes from NE India (secondary centre of origin for chilli), 58% occurred among individuals within populations (Yumnam et al. 2012). This study also showed that based on molecular data, type specimens of *C. annuum*, *C. chinense* and *C. frutescens* species clustered together. Additionally, understanding the mechanism(s) underlying pungency (one of the most important traits of commercial value and a crucial component of flavour) is still emerging (von Steimker et al. 2024). An untargeted approach will lead to a better understanding of the biochemical and molecular mechanisms underlying this metabolic variability across species and can contribute towards breeding programs to develop new commercial hybrids with fruits of the desired characteristics. We also tried to identify key biosynthetic pathway genes and associated allelic variation. This study, along with previously reported allelic variation by several workers for both colour and flavour suggests that some of the metabolites identified in this study can be used as biomarkers for further characterisation of these heirloom chilli and could provide additional parameter(s) in distinguishing improved cultivars from landraces. The underlying molecular mechanism for the health-related compounds once understood better, will also contribute to breeding chillies enriched in health-related compounds.

Materials and methods

Plant material and experimental design

Chilli genotypes include *Capsicum annum* [Dalle khursani (AL-2), J-41(B) and Kashi Anmol (KA)], *Capsicum chinense* [King chilli/Umorok] (AL-1) and *Capsicum frutescens* [Sohmynken khnai] (AL-3). Seeds from ripe fruit were sown in seedling trays containing a mixture of soil, coco-peat and decomposed farmyard manure in a ratio of 2:1:1. Matured seedlings were transplanted in pots in three replications and were kept in a fiber house, maintaining moderate moisture supply, temperature and humidity. This experiment was conducted in the College of Post Graduate Studies in Agricultural Sciences, Umiam, Meghalaya, in February-June 2021. Mature and ripe whole chilli samples (five fruits per plant and a total of three plants per genotype) were collected and kept for oven drying (50–55 °C for 48 h). Dried whole fruit samples of different plants belonging to the same genotype (n = 15) were pooled and divided into three aliquots to assess the technical reproducibility of metabolite extraction in fruit samples.

Metabolite extraction and LC-high-resolution MS (LC-HRMS) analysis

Dried chilli fruit samples (100 mg) were homogenized in liquid nitrogen and subsequently extracted with 800 µL of ice-cold 80% methanol containing 0.1% formic acid. The mixture was vortexed, shaken, and sonicated before centrifugation. The supernatant was evaporated, and the residue was re-suspended in 200 µL of the same solvent and filtered. The filtered extract (50 µL) was loaded into an auto-sampler vial, which was then placed in an ultra-high performance liquid chromatography (UHPLC) auto-sampler at 10 °C. For untargeted plant metabolomics, we followed the protocol described by Avuthu et al. (2024). The analysis was conducted using UHPLC-quadrupole time-of-flight mass spectrometry (UHPLC-Q-TOF-MS/MS) in positive ionization mode (electrospray ionization, ESI) with a Xevo-G2-XS mass spectrometer (Waters, Milford, MA, USA). In this mode, the capillary and cone voltages were set to 2.0 kV and 40 V, respectively. The desolvation temperature was maintained at 400 °C, the source temperature at 110 °C, with desolvation gas flow at 800 L/h and cone gas flow at 50 L/h. Scanning was performed with a time of 1 s and a mass range of 50–2000 Da. Chromatographic separation was achieved using an acquity UPLC BEH-C18 column (100 mm × 2.1 mm, 1.7 µM particle diameter, Waters) maintained at 30 °C. The mobile phases were 0.3 mL/min of ultra-pure water with 0.1% formic acid (mobile phase A) and methanol with 0.1% formic acid (mobile phase B). The elution gradient was as follows: 0–2 min at 1% B; 2–6 min at 20% B; 6–10 min at 60% B; 10–15 min at 90% B; 25–31 min at 5% B; 31–35 min at 1% B. A multiplexed MS/MS method with alternating low and high energy (MSe) in centroid mode was used for data acquisition. Mass Lynx 4.1 software (Waters) was used for data analysis, and leucine encephalin (m/z 556.2771, [M-H]+) was applied for lock mass calibration as previously reported (Darko et al. 2022). In brief, leucine encephalin, was continuously introduced into the mass spectrometer during analysis. This process corrects any drift in the mass axis, ensuring precise mass calibration throughout the run and enabling consistent data comparison across multiple runs. Additionally, blank samples—devoid of analytes—were analyzed to confirm the absence of contamination or carryover between samples. Calibration standards, such as sodium iodide, with known concentrations, were used to verify the system's quantitative accuracy. Additionally, matrix effect assessment was performed using a dilution series, starting with undiluted samples and dilution in 1:2, 1:4, 1:8, and 1:12 series (Supplementary Fig. S1). This stepwise approach allowed for the determination of proportional response drops, a key indicator of matrix effects. The experiment was carried out using two distinct genotypes (J-41(B) and KA) to evaluate whether matrix effects were consistent across different samples.

The LC-HRMS output files were converted to mzXML format and processed using software MZmine-2.53 (Pluskal et al. 2010). The data, including observed masses, m/z values, retention times, and relative intensities, were then imported into MS Excel. Peaks that were inconsistent across replicates, as well as those identified as isotopes and adducts, were excluded from further analysis.

Metabolomic data analysis

MS excel containing a list of compounds with peak height data were used for data analysis. The metabolites were putatively identified based on an accurate mass match (accurate mass error (AME) < 10 ppm) (Summer et al. 2007), with metabolites reported in different databases: METLIN, Plant Metabolic Network (PMN), LIPIDMAPS, METLIN, and KEGG. The data was log-transformed and normalized using median values and uploaded to the MetaboAnalyst 5.0 [https://www.metaboanalyst.ca] (Pang et al. 2021) for principal component analysis (PCA) and partial least square discriminate analysis (PLS-DA) to observe the overall distribution between samples (Supplementary Table S1).

The previously reported metabolites from capsaicinoid pathway (capsaicin, homocapsaicin, dihydrocapsaicin, capsorubin, capsanthin 3,6-epoxide and capsi-amide) were checked for their retention time and peak intensities across different runs in the chilli genotypes (Supplementary Fig. S2). Capsaicinoid complex metabolites were considered as they are unique to chilli and of commercial importance. Our analysis consistently detected these key metabolites across all collected data. The retention times of these compounds remained uniform across all analyzed samples (Supplementary Fig. S2), demonstrating the consistency and reliability of the obtained results. This approach to confirm data robustness is similar to reported targeted metabolite profiling (Darko et al. 2022).

Differentially accumulated metabolites were selected using the criteria of $P < 0.05$ and $\log_2 FC > 1.0$. Volcano plots were used to visualize, and screen differential metabolites based on P value and fold change value. The volcano plots were generated using SR plot online (https://www.bioinformatics.com.cn/en?keywords=volcano). Venn diagrams were generated using the Venn Diagram (https://bioinformatics.psb.ugent.be/webtools/Venn/). The heatmap was generated using the TB-tools v2.0 (Chen et al. 2020).

Capsaicin estimation and allele mining across key biosynthetic genes of the pathways identified

For capsaicin estimation, the standard curve of pure capsaicin was prepared using capsaicin (RM10095, HiMedia,

India) and ethanol (Absolute, 99% purity, ACS grade) as solvent. 10 mg standard capsaicin was taken in a 100 ml volumetric flask. 50 ml ethanol was added and stirred well using a magnetic stirrer for 10 min then diluted by ethanol up to the mark of the volumetric flask to prepare the final strength of the solution (100 µg/ml). The estimation was done at 280 nm using an MS UV Plus Spectrophotometer (Motras Scientific, India) and revealed a standard curve with $y = 0.0077x + 0.0045$ with a $R^2 = 0.9981$. This standard was used for estimating the capsaicin content of the five chilli genotypes.

The mature chilli fruits were harvested (five fruits/plant and a total of three plants/genotype) and oven-dried for 8 h at 45 °C. The dried fruits were ground to a fine powder using mortar and pestle. 500 mg of ground chilli powder was incubated in 5 ml of ethanol for 24 h. An additional 2 mg of activated charcoal was added to reduce of coloured pigment. Samples were filtered using a 0.45 µm syringe filter. The filtered samples were analysed using MS UV Plus (at 280 nm) and capsaicin content was estimated using the above-mentioned standard curve. The formula for calculating Scoville heat units is as follows SHU = capsaicin concentration (in ppm) × 15,000.

The capsaicin content identified by the untargeted approach was cross-checked with the values obtained using the commercial capsaicin for validation. The values derived from the reference standard were compared with average values obtained from the exact mass (with AME < 10 ppm) of metabolite IDs C06866 (capsaicin) and CPD9185 (dihydrocapsaicin) and a sum of C06866, CPD9185 and C20215 (homocapsaicin) for the five genotypes.

Allele mining involved the identification of sequences for the targeted genes followed by examination of sequence variation. Briefly, the gene name was used as input in the search option for downloading the protein or CDS sequence from the SGN (Sol Genomics Network) database. This sequence was used for performing BLAST in the pepper genomics database (pepper GD) and gene IDs obtained. These gene IDs were checked for gene ontology terms. Only IDs having the reported/characteristic InterPro domains were selected for further analysis. The variants (SNPs/small InDels) across the three species *Capsicum annuum*, *C. frutescence* and C. *chinense* were identified. The type of variants viz., UTR variants, splice variants, intron, synonymous, missense, gain or loss of start or stop codons and InDel were identified. The frequency of alternate alleles across the three species for putative functional polymorphism was calculated.

Results

Variation in the fruit morphology

The four landraces, when compared with Kashi Anmol (KA) using the guidelines provided by Protection of Plant Varieties and Farmers' Rights Authority (PPV&FRA, GoI) for chilli (hot pepper), bell (sweet pepper) and paprika (*Capsicum annuum* L.) (Gazette of India 2015), have morphologically unique traits (Table 1; Fig. 1). *C. chinense* (Bhoot Jolokia/Naga chilli/Umorok) (AL-1) fruit is moderately triangular with broad base varying from sub-conical to conical shape having glossy irregular rough skin. *C. annuum* (Dalle

Table 1 Morphological details regarding fruit shape, size and colour are based on guidelines for chilli given by PPV&FRA [17] for the genotypes used in the current study. The data is the average of five fruits per three plants for each genotype (n = 15)

Genotype (s)	Fruit shape	Fruit size		Fruit colour	
		Length (cm)	Diameter (cm)	Matured unripe	Matured ripe
Capsicum chinense cv. Umorok (AL-1)	Moderately triangular with a broad base and tapering tip Or Subconical to conical with straight to low curvature apex (medium to high glossy irregular rough skin)	4.9–5.5	3.2–3.5	Green (light to medium)	Red (medium to dark)
Capsicum annuum cv. Dalle khursani (AL-2)	Cordate (with glossy and smooth texture)	1.6–1.8	1.7–1.8	Green (Medium to Dark)	Red (light to medium)
Capsicum frutescens cv. Sohmynken khnai (AL-3)	Narrowly triangular (smooth and medium glossy skin)	3–3.4	0.8–1.1	Green (light to medium)	Red (light to medium)
Capsicum annuum [J-41(B)]	Moderately triangular (glossy smooth pericarp)	4.6–4.9	0.6–1.1	Green (medium to dark)	Red (dark)
Capsicum annuum cv. Kashi Anmol (KA)	Moderately triangular (glossy smooth pericarp with low curvature)	5.6–6.4	0.6–0.9	Green (dark)	Red (dark)

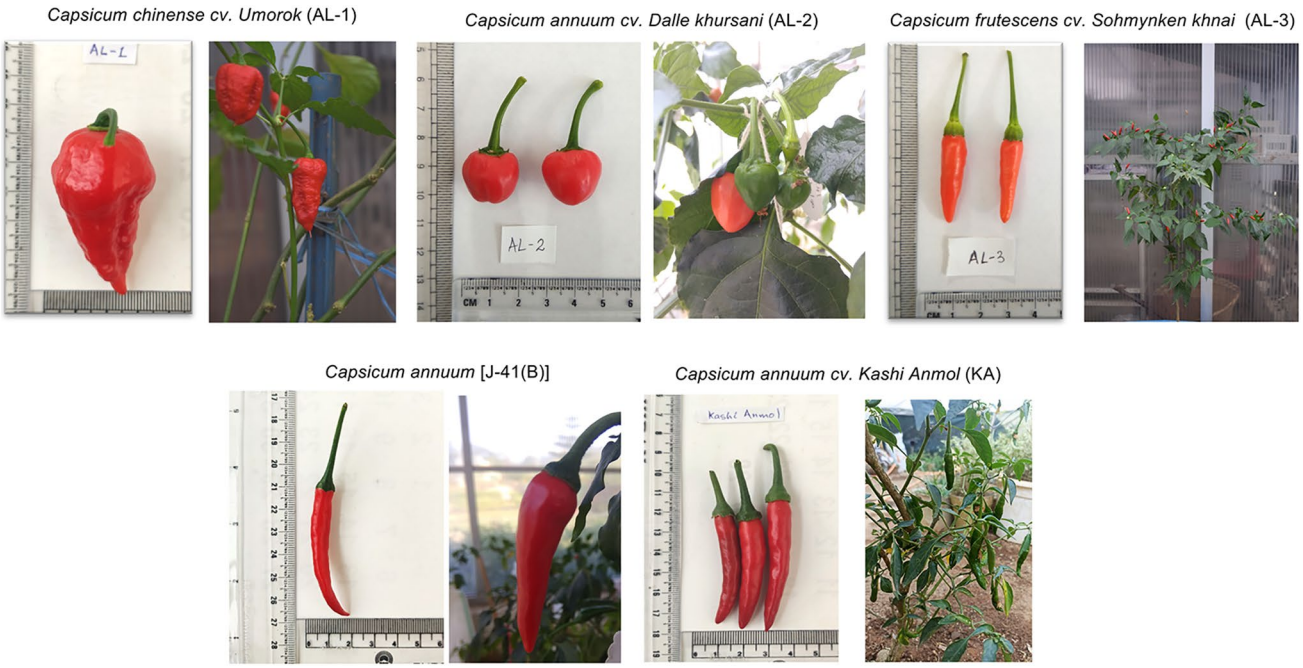

Fig. 1 Representative picture showing morphological variation in the mature fruits of five chilli genotypes. The four heirloom landraces from NE India (AL-1, AL-2, AL-3, and J-41(B) were compared with KA. AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is a *chinense* and AL-3 is a *frutescens*

khursani) (AL-2) fruit is cordate in shape with glossy and smooth skin; *C. frutescens* (Sohmynken khnai) (AL-3) is narrowly triangular chilli having smooth and medium glossy skin with 3–3.4 × 0.8–1.1 cm fruit dimension, which is relatively a smaller chilli size than other species. *C. annuum*, such as J-41(B) and Kashi Anmol (KA), are moderately triangular and have glossy smooth skin with low curvature. The fruit shape index (ratio of length to diameter) for the genotypes was 1.64, 1.03, 3.58, 5.76 and 8.53 for AL-1, AL-2, AL-3, J-41(B) and KA, respectively.

Differential accumulation of metabolites in chilli genotypes

Untargeted metabolomic analysis was performed to understand the biochemical changes in the different species of chilli genotypes. A total of 6990 consistent peaks of monoisotopic masses were detected in all the replicates, out of which 2702 metabolites were identified using accurate mass error < 10 ppm in MZmine 2.53 software (details about methodology available in materials and method section (Supplementary Tables S1 and S2)).

A multivariant analysis was used to check the robustness and repetition of our experimental design (loading file used for PCA available as Supplementary Table S1). A Principal Component Analysis (PCA) score plot was constructed to elucidate the similarities and distinctions between genotypes, offering insights into potential sources of variability or bias in the data using Metaboanalyst software (Fig. 2).

The first principal component (PC1) accounted for 63% of the total variance, effectively discriminating the genotypes AL-1, AL-2, and AL-3 with KA and J-41 (B). The second principal component (PC2), explaining 14.4% of the variance, distinguished KA with J-41 (B). However, no differentiation was observed for genotypes AL-1, AL-2 and AL-3 (Fig. 2a). Further, partial least square discriminate analysis (PLS-DA) component 1 clearly differentiated genotypes AL-1, AL-2 and AL-3 from genotypes KA and J-41(B) with a variation of 62.9% and component 2 separated genotypes AL-1, AL-2 and AL-3 explaining a variation of 7.4% (Fig. 2b). The clear separation underscored the reliability of results and highlighted the distinct nature of genotypes. The analysis clearly showed differences among the five genotypes with treatments within genotypes clustering together. The analysis also suggests that commercial genotype KA is distinct from all the other genotypes from the NE region, including J-41(B), which belongs to the same species and has similar fruit characteristics. The genetically diverse *C. annuum* genotypes can be crossed for obtaining heterosis for traits of commercial importance. It is known that the *C. annuum* complex, including the three species, underwent selection several times for key commercial traits. Out of *C. annuum* species samples only AL-2 clusters with *C. chinensis* and *C. frutescence*. As the clustering is based on metabolites, it is possible that AL-1 and AL-2 landraces

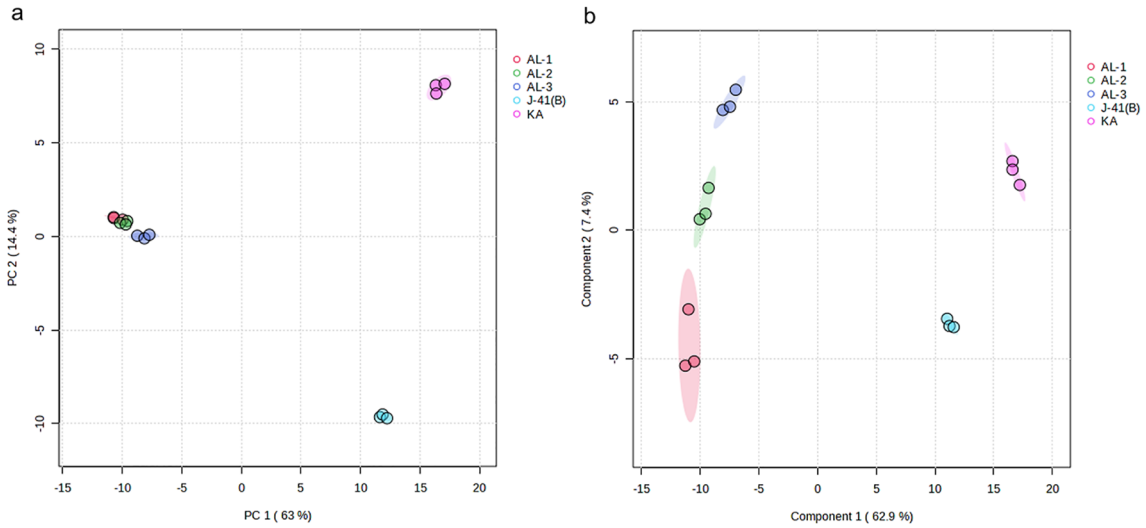

Fig. 2 Multivariant analysis of significant metabolites from mature fruits of five chilli genotypes using PCA (**a**), and PLS-DA (**b**). The significant metabolites ($P < 0.05$) from chilli genotypes form separate groups, indicating an altered metabolite level state in the fruit. The genotypes are represented with distinct colours, as indicated in the top right corner of the figure. AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is a *chinense* and AL-3 is a *frutescens*

were selected for similar fruit quality traits during domestication as revealed by overlapping of samples (Fig. 2a) despite having very distinct fruit shapes.

Metabolomic analysis was performed to understand the metabolomic changes in the *C. annuum* (J-41(B), AL-2 and KA), *C. chinense* (AL-1) and *C. frutescens* (AL-3). The mature fruits of genotypes J-41(B) and KA were deep red compared to other genotypes. A total of 1704 metabolites that were differentially accumulated were identified in all combinations (Supplementary Table S2). A higher number of differentially accumulated metabolites were seen in the J-41(B) versus AL-3 (1376), followed by J-41(B) versus AL-2 (1365), J-41(B) versus AL-1 (1257), KA versus AL-2 (649), AL-3 versus KA (616), KA versus AL-1 (594) and J-41(B) versus KA (413) (Fig. 3), suggesting that variation between species was higher than within species. Volcano plots were generated to identify significantly upregulated and downregulated metabolites in all possible pair-wise comparisons (Fig. 3). In addition, a higher number of upregulated metabolites were observed in AL-1 versus JL-41(B) (↑859 and ↓485). However, the number of downregulated metabolites was higher in J-41(B) versus AL-3 (↑463 and ↓933) comparison. Differentially accumulated metabolites were further explored to identify biochemical pathways using pathway enrichment analysis.

Pathway enrichment analysis

Metabolite pathways were determined by conducting a KEGG mapper analysis, considering all possible pair-wise comparisons. The top three pathways were identified based on their significance and involvement in the observed metabolic changes. The outcomes were employed to construct pathway visualizations.

Species-specific pathway enrichment shows several metabolites associated with diterpenoid biosynthesis, arachidonic acid metabolism, and biosynthesis of unsaturated fatty acids were significantly enriched (Supplementary Fig. S3).

Diterpenoid biosynthesis is a significantly enriched pathway observed across all the species-specific pairwise comparisons (i.e. *Capsicum annuum* vs *Capsicum annuum, Capsicum annuum* vs *Capsicum chinensis, Capsicum annuum* vs *Capsicum frutescens* and *Capsicum chinensis* vs *Capsicum frutescens*) (Supplementary Fig. S3). However, unsaturated fatty acids (including α-linoleic acid) were significantly enriched in *Capsicum annuum* versus *Capsicum chinensis, Capsicum annuum* versus *Capsicum frutescens and Capsicum chinensis* versus *Capsicum frutescens* comparisons only. Further, cutin, suberin, and wax biosynthesis were enriched in *Capsicum annuum* versus *Capsicum chinensis*, *Capsicum annuum* versus *Capsicum frutescens* and *Capsicum chinensis* versus *Capsicum frutescens* comparisons but not in *Capsicum annuum* versus *Capsicum annuum* (Supplementary Fig. S3).

Identification of species-specific metabolites in chilli genotypes

A total of nine major categories of metabolites viz, fatty acyls, sterol lipids, flavonoids, alkaloids, biosynthesis of secondary metabolites, terpenoids, phytochemical

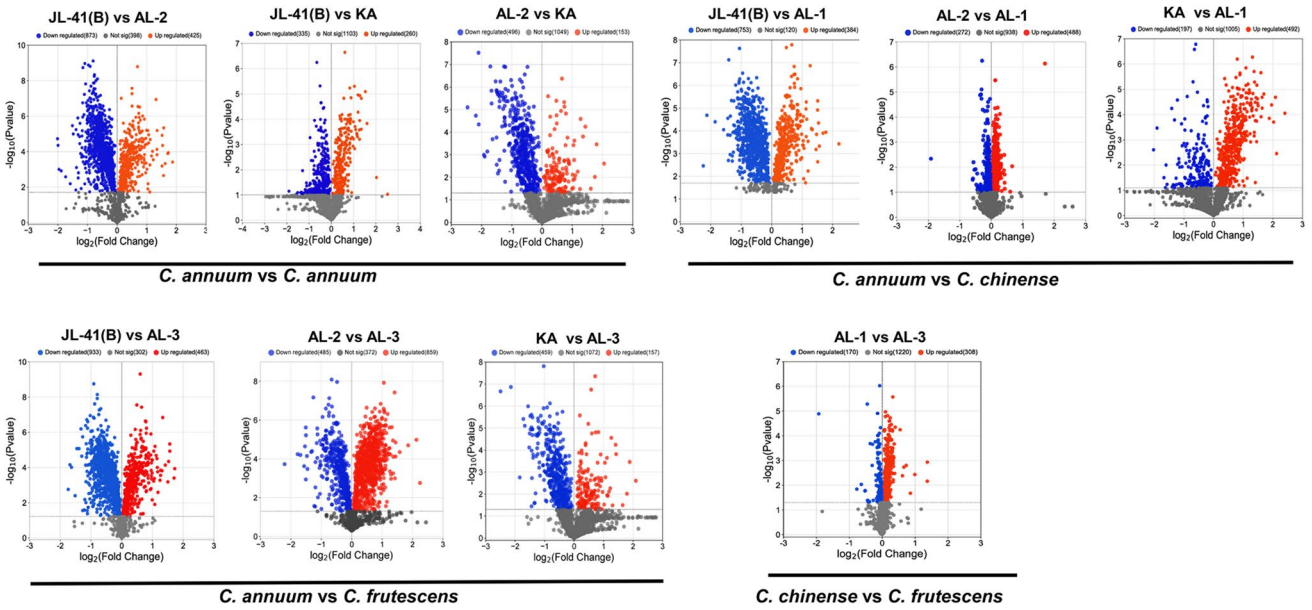

Fig. 3 Identification of differentially accumulated metabolites in mature chilli fruits using pair-wise genotypic comparisons. Volcano plots were generated based on significantly accumulated metabolites using criteria $P < 0.05$ and $\log_2$ fold change > 1. In these plots, red and blue indicate up- and down-regulated metabolites, respectively. AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is a *C. chinense* and AL-3 is a *C. frutescens*. Details of metabolites up- or down-regulated are available in Supplementary Table S2

compounds, glycosides and other metabolites were identified (Fig. 4). The actual composition (in percentage) varied in different pairwise comparisons. The three major categories, fatty acids, sterol lipids and flavonoids, contributed to more than 70% of the significantly expressed metabolites across genotypes.

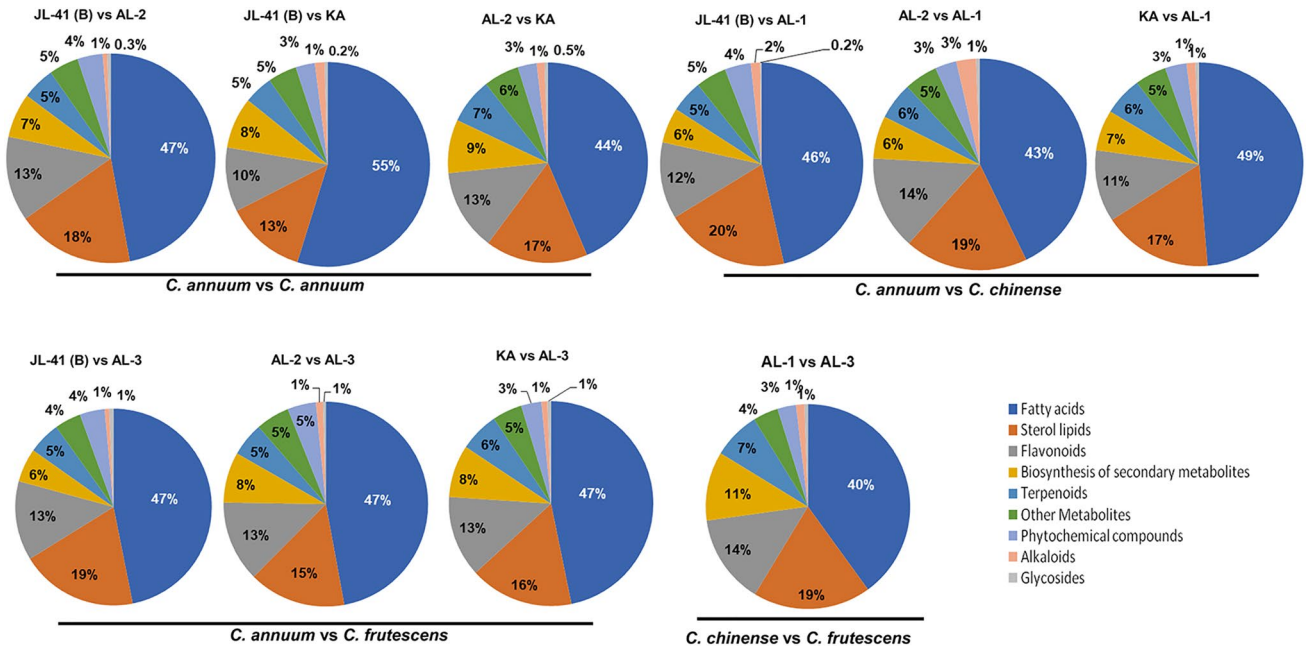

Fig. 4 Percent distribution of major categories of significantly accumulated metabolites in mature fruits of chilli genotypes. Significantly accumulated metabolites were identified using criteria $P < 0.05$ and $\log_2$ fold change > 1 and pathways were identified using the KEGG mapper. Pair-wise genotypic comparison is depicted for nine major categories of metabolites identified. AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is a *C. chinense* and AL-3 is a *C. frutescens*

Differentially accumulated metabolites selected using the criteria of $P < 0.05$ and $\log_2 FC > 1.0$ were used to generate a heatmap using online TB tools that led to the identification of five distinct genotypic clusters (Fig. 5). The first cluster suggests that high fold change carotenoids and flavonoids can be targeted for improving the colour and flavour in the three landraces (AL-1, AL-2 and AL-3). While other highly accumulated metabolites underlying colour and flavour were identified for cluster two. In the case of cluster three, flavonoid and terpenoid were identified as the key metabolic categories. For the other two clusters, no unique cluster specific pattern could be obtained. This suggests that there is enough variation within clusters four and five to be used for breeding fruit traits of importance.

The significant high-fold metabolites from the heatmap belong to four major biosynthetic pathways. Some of the important metabolites for each of the pathways are as follows.

Carotenoid pathway

Compounds of carotenoid pathway like monoanhydrobacterioruberin (3.93-fold), isorenieracistene (9.37-fold), hydroxyspirilloxanthin (2.05-fold) and 2-ketospirilloxanthin (1.88-fold) are present in higher concentration in AL-1 compared to J-41(B) (Fig. 5). On the other hand, compounds like rhodovibrin (2.12-fold), phoenicoxanthin (2.68-fold) and 2,2'-diketospirilloxanthin (1.58-fold) are present in higher concentration in J-41(B) compared to AL-1. Compounds 3,4-dihydroanhydrorhodovibrin (1.75-fold), 3',4'-dihydrorhodovibrin (1.98-fold), adonixanthin (4.48-fold), canthaxanthin (1.30-fold), phoenicoxanthin (2.69-fold) and rhodovibrin (2.38-fold) are higher in J-41(B) compared to AL-2. On the other adonixanthin (5.81-fold), 2,2'-diketospirilloxanthin (1.52-fold), 3,4-dihydroanhydrorhodovibrin (1.57-fold), 3',4'-dihydrorhodovibrin (1.89-fold), canthaxanthin (1.30-fold), phoenicoxanthin (2.86-fold) and rhodovibrin (3.11-fold) are higher in J-41(B) compared to AL-3.

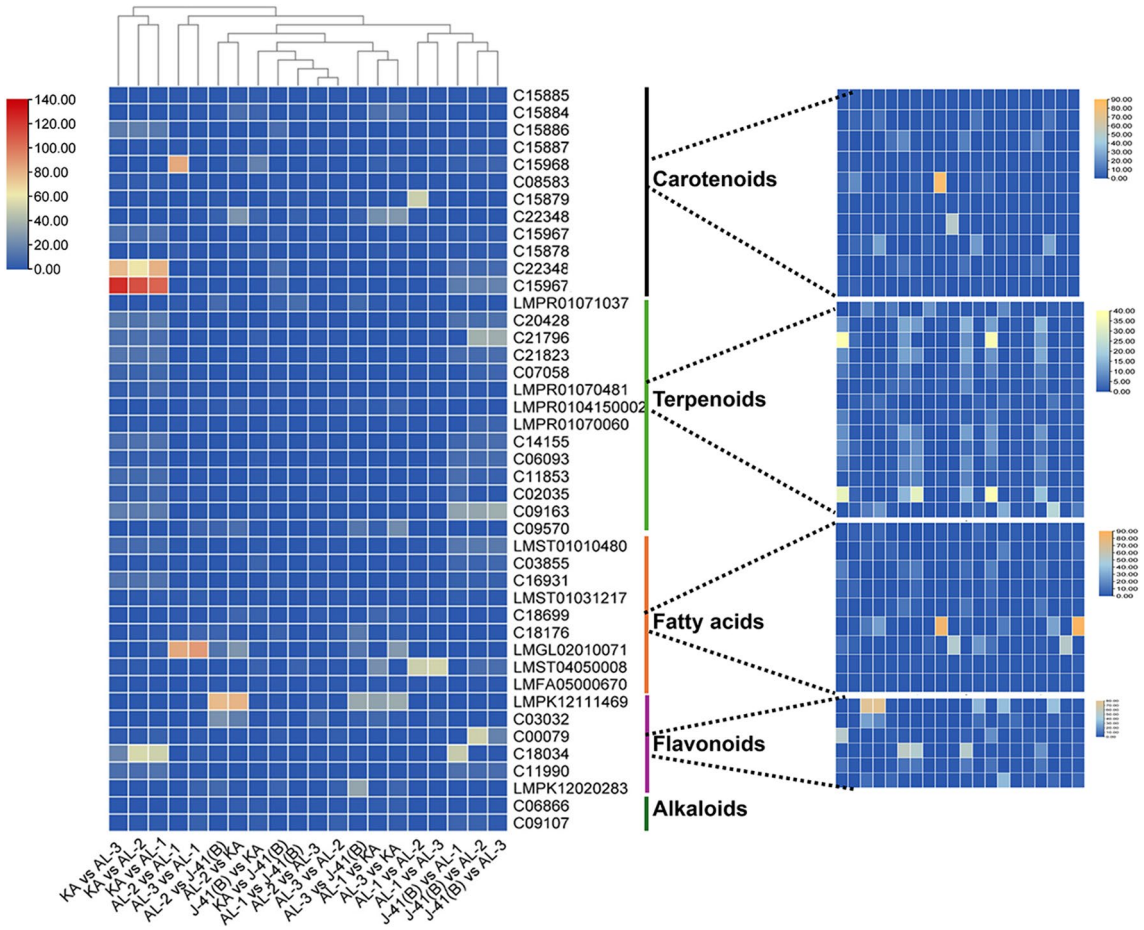

Fig. 5 Heat map analysis of the differentially accumulated metabolites observed in chilli genotypes. The x axes indicate pairwise comparison while y axes are KEGG identifiers for metabolite as mentioned in Supplementary Table S2. The red and blue colour represents the increased or decreased accumulation of metabolites, respectively. A zoomed in version of the heatmap for the four major classes is shown on the right side

Compounds like 3,4-dihydroanhydrorhodovibrin (17.22-fold), canthaxanthin (2.92-fold), phoenicoxanthin (9.91-fold) and hydroxyspirilloxanthin (1.19-fold) are higher in KA compared with AL-2. There was a 6.78-fold higher amount of jasmonyl-L-leucine in genotype AL-1 as compared with KA.

Fatty acid pathway

The levels of various glycerophospholipids were different in the five genotypes. PA (17:0/0:0) and PA (20:5(5Z,8Z,11Z ,14Z,14Z,17Z)/22:0) content were 2- and fourfold, respectively higher in AL-1 in comparison with KA. Dinorursodeoxycholic acid content was 24-fold higher in AL-1. PC (2:0/2:0) content in AL-1 was 45-fold higher (KA). DGTS (16:0/18:3(6Z,9Z,12Z)) content was also substantially (19-fold) more in AL-1. AL-1 fruit has a rough texture compared to other genotypes evaluated in the present study. Dinorursodeoxycholic acid, belonging to the main class 'bile acid and derivative', is present 24-fold higher in AL-1 compared with KA. Cinobufagin (sterol lipid; bufanolide) is 11-fold more in KA as compared to AL-1. J-41(B) has higher amounts of steroids like certonardosterol G (13.6-fold) and cholesterol-beta D glucoside (6.8-fold more) in comparison to AL-1. AL-2 has higher levels of ehrensteroid C (an ergosterol). 5S-hydroperoxy-18R-HEPE, an eicosanoid is present in six-fold higher amounts in AL-3 compared with AL-1. Additionally, one of glycerolipids (DG (18:2(9Z,12Z)/18:3(9Z ,12Z,15Z)/0:0) [iso2]) is present at nearly 88-fold higher in AL-3. In AL-1, many sterol lipids were identified but all diverged to either degradation or primary bile derivatives. Content of two prenol lipids (tutin 22 times more and 3β-(3-methyl-butanoyloxy)-villanovane-13α,17-diol at 25-fold) were significantly higher in AL-3 (cf KA). Prenol lipids- tephrowatsin B, 3,4,3',4'-tetrahydrospirilloxanthin-20-al, 3β-(3-methyl-butanoyloxy)-villanovane-13α,17-diol and tutin were found in higher amount in AL-3 (vs AL-1).

Flavonoid pathway

All the chilli genotypes analysed in the present study are pungent with a distinct flavour. In KA and J-41(B), a higher number of isoprenoids (C40 type; like echinenone) were found. In KA, nearly fourfold higher levels of L-phenylalanine as compared to AL-1 were noted. Dinorursodeoxycholic acid content in AL-1 is 24-fold more (cf KA). Additionally, a 6.7-fold higher amount of jasmonyl-L-leucine in AL-1 was detected in comparison with KA. Prenol lipids such as (3S)-11-cis-3-hydroxyretinal, 13-cis-retinoic acid and phorbol 12-tiglate 13-decanoate were present in significantly higher amounts in fruits of KA. Secondary metabolites such as 11-hydroxyferruginol, 11-hydroxysugiol, gibberellin A14 aldehyde and pisiferic acid were also present in higher amounts. A flavanone-malonylapiin was also in very large amounts in KA. Polyketide-5,7,8,3',4'-pentahydroxy-6-methoxyflavone content in AL-2 was 79-fold more than KA. Cyclopiazonic acid was also present at ninefold higher levels. The capsaicin content in AL-3 was threefold higher compared to KA. Staphylopine (a type of D amino acid) was twofold higher in AL-3 compared with KA.

Secondary metabolites like 11-hydroxyferruginol, 11-hydroxysugiol, catharanthine, dTDP-D-forosamine, γ-tocotrienol, gibberellin A12 aldehyde, gibberellin A14 aldehyde, gibberellin A20, oleandolide and pisiferic acid were higher in J-41(B) versus AL-1.

Based on the metabolites obtained, we were able to identify fatty acid metabolism with phosphatidic acid leading to variation in the distribution of glycerophospholipids, sphingolipids and steroids in the cell membrane along with carotenoid, flavonoid and capsaicinoid pathways as key pathways differentiating chilli accessions (Fig. 6).

Capsaicin estimation and allele mining of key biosynthetic genes across the three *Capsicum* species

Capsaicin content using commercially available standards was estimated for the five chilli genotypes. The capsaicin content ranged from 7.4 ppm (111,023.8 Scoville Heat Units (SHU)), 5.6 ppm (84,113.77 SHU), 5.38 ppm (80,828.77 SHU), 4.47 ppm (67,178.77 SHU) and 3.45 ppm (51,848.77 SHU), respectively for the genotypes AL-1, AL-2, AL-3, J-41(B) and KA. The estimated capsaicin values showed a positive correlation of 0.46, 0.71 and 0.61 with capsaicin, dihydrocapsaicin and total capsaicinoid content, respectively.

A set of 21 genes belonging to flavonoid, capsaicinoid and carotenoid pathways (Fig. 6; Table 2) were explored to see whether species-specific alleles exist which could be targeted for breeding programs. The metabolomic approach led to the identification of key biosynthetic pathways which in turn resulted in identifying the underlying genes using standard ontology-based methods using both Solanaceae and Pepper databases (details available in the Methods section). This was followed by allele mining using the pangenome data available for a set of 326 diverse genotypes (CA-112; CF-99; CC-115). For most of the biosynthetic genes, alternate alleles exist (leading to missense, stop, gain of start codon) which could be putative targets for designing markers underlying such potential functional nucleotide polymorphism (FNP). The frequency of alternate alleles ranged up to 100% for enzymes like KAS and CAT in the case of CC. For some of the isoforms of *pAMT* (03g05690 and 03g22890), there is no variation reported in the pangenome.

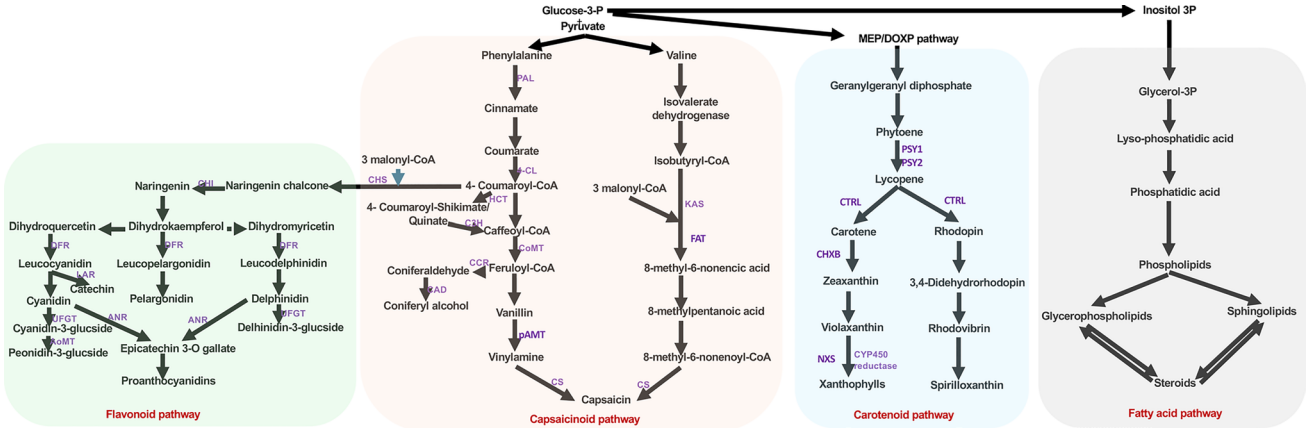

Fig. 6 Secondary metabolic pathway shows the interactions of different biochemical compounds in mature chilli fruits. AL-2, J-41(B) and KA belong to *C. annuum* species, AL-1 is a *chinense* and AL-3 is a *frutescens*. The key biosynthetic genes used for allele mining are indicated in purple colour

For the seven genes analysed for the flavonoid pathway, no stop codon was identified (Table 2). The total number of variants ranged from 8 to 65. Stop codon was identified in five genes.

Discussion

Chillies are recognised as one of the most important spices. The molecular basis of quality traits, including flavour and other fruit attributes, are still not properly understood. Multiple studies on diversity in chilli accession (Yumnam et al. 2012; Rivera et al. 2016; Du et al. 2019; López Castilla et al. 2019; Paredes Andrade et al. 2020) and some of the domestication-related traits including fruiting behaviour and fruit size have been conducted (González-Pérez et al. 2014; Du et al. 2019; Liu et al. 2023). Despite the commercial and health benefits of chilli, limited research has been conducted to characterise and understand the extent of variation in metabolites in chilli and their association with key fruit traits like shape, colour, texture and flavour. A previous analysis of GC–MS-based metabolite profiling for 136 chilli genotypes identified only 56 metabolites (Gaur et al. 2016). A similar approach was also used to study the accumulation of capsacinoids in inter- and intra-specific hybrids of two species of *Capsicum* peppers, *C. chinense* (cv. Habanero and cv. Biquinho) and *C. annuum* var. *annuum* (cv. Jalapeño and cv. Cascadura Ikeda) (Naves et al. 2022). This study suggested that heterosis for capsaicinoids is dependent on the parent-of-origin effect. Targeted metabolite approach for both capsaicin pathway-related metabolites and antioxidants is well-reported in chilli (Reddy et al. 2014; Das et al. 2021; Lahbib et al. 2021; Akhter et al. 2024; Naseem et al. 2024; von Steimker et al. 2024). Using an untargeted metabolite approach, we could identify 2702 putative metabolites, across chilli species. This extensive variation in metabolites was identified for the four-heirloom chilli of NE India compared with Kashi Anmol (KA). The genotype AL-1 or Naga chilli/bhut jolokia is a native *Capsicum* landrace found in NE India, which is one of the naturally occurring hottest chilli in the world. From a molecular breeding perspective, a better understanding of the key fruit traits will allow for better-designed crop improvement strategies (Dutta et al. 2018; Peñuela et al. 2021). Previously, we had identified morpho-molecular markers for diverse accessions from NE. We also identified variations in the pungency related locus *Pun1* (Yumnam et al. 2012). The current study was able to identify genotype-specific metabolites including variation in capsaicin. Our findings validate that all the genotypes used in the present study are pungent while the commercial genotype KA has the lowest pungency levels. Additionally, the differential accumulation of metabolites belonging to the carotenoid pathway was also identified. The rationale for the selection of these four different chilli types was to identify metabolites which would help generate a baseline dataset for further studies on fruit quality traits in chilli.

Previously, metabolomic profiles during fruit development in domesticated and wild accessions of chilli fruits (*C. annuum*) revealed that the domesticated group had the highest level of glycerophospholipids, with an abundance of lipid metabolism, a decrease of capsaicinoids, terpenoids, flavones, and flavanol compounds when compared with their wild ancestors (Cervantes-Hernández et al. 2022). Population structure using 467 diverse chilli accessions has also revealed that the levels of diversity negatively correlate to levels of domestication, with the more diverse being the least domesticated (McCoy et al. 2023). Our analysis of the four diverse heirloom chilli landraces from NE India suggests that the diversity of metabolites is still present, which needs to be harnessed. The metabolomic approach led to

Table 2 Types of variants available across twenty-one genes targeting three biosynthetic pathways using a set of 326 chilli genotypes. CA, CF and CC refer to *Capsicum annuum, Capsicum frutescens* and *Capsicum chinensis*, respectively

Gene name	locus Id	Gene size (kb)	Types of variants (SNP + InDel)											% alternate allele (CA allele is taken as reference allele)			Putative FNP (Total)
			5'UTR	3'UTR	Splice variant	Intron	Synonymous	5'UTR premature start codon	Missense & splice variant	Missense	Stop gained	InDel	Total	CA (112) (%)	CF (99) (%)	CC (115) (%)	
CHS	05g18580	2.44	4	9	0	27	7	1	0	3	0	0	52	1	2	87	1 (4)
	03g38299	1.84	3	0	0	17	16	0	0	4	0	0	40	0	64–72	20–80	3 (4)
CHI	02g08200	1.1	0	0	3	3	4	0	0	3	0	0	13	0	11	18	1 (3)
	06g00690	4.9	3	2	0	33	4	1	0	2	0	9	54	0	55	97	1 (4)
DFR	02g23230	1.69	0	0	2	16	4	0	0	7	0	0	29	0	20–94	99–100	4 (7)
LAR	03g38730	1.14	0	0	0	0	10	0	0	4	0	0	14	1	7–76	64–85	4 (4)
ANR	04g02940	1.98	0	0	1	18	12	0	1	7	0	0	39	0	1–99	36–100	5 (8)
UFGT	10g15690	1.34	0	0	0	0	2	0	0	6	0	0	8	0	31	1	1 (6)
AOMT	08g21450	2.25	1	7	1	43	7	1	2	3	0	0	65	0	29	98	1 (6)
PAL	09g20190	4.1	3	2	1	17	10	1	0	5	0	0	39	1	2–48	3–89	5 (6)
	09g20211	2.36	0	0	0	0	2	0	0	2	0	0	4	1	98	100	1 (2)
	10g10640	4.25	8	3	0	66	22	1	0	7	0	0	107	1	28–92	1–100	6 (7)
4CL	06g11550	4.94	4	8	2	102	22	2	0	8	0	21	169	0	1	53	1 (10)
HCT	03g05670	5.67	4	12	0	133	19	2	0	11	1	0	182	1–6	12–75	2–77	8 (14)
KAS	02g04370	6.17	3	3	0	198	10	0	0	12	0	19	245	1–7	1–99	1–00	2 (12
	10g05730	2.64	17	3	1	17	11	1	0	7	0	5	62	0	26–54	20–45	2 (8)
FAT	06g01810	6.95	0	0	2	88	9	0	0	5	0	8	112	22	5–90	97–100	3 (5)
C3H	01g17970	10.35	0	0	0	281	6	0	0	14	1	26	328	0	3–99	1–100	9 (17)
	08g09910	2.93	3	7	1	15	14	0	0	5	0	12	57	0	49–52	2–23	3 (5)
CCoAMT	09g16200	1.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCR	01g16400	10.46	0	0	4	533	4	0	0	25	0	31	597	0	2–96	7–97	13 (29)
	01g35570	0.6	0	0	0	21	5	0	0	8	0	4	38	0	1–54	3–97	5 (8)
pAMT	01g24960	1.07	0	0	0	0	8	0	0	15	3	0	26	1–2	1–95	1–95	18 (18)
	01g26200	7.32	6	1	0	190	9	1	0	4	0	26	237	1	21–96	6–100	7 (10)
	03g05690	1.47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	03g22890	3.71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	07g16400	1.16	0	0	0	9	3	0	0	13	1	6	32	1–89	20–89	1–100	9 (15)
	12g08780	4.2	0	0	3	111	6	0	0	10	0	24	154	1–3	2–100	3–100	7 (11)
	12g08790	6.34	4	10	2	177	6	0	0	18	0	25	242	1–5	3–100	4–100	13 (18)
CS	02g22600	3.38	0	0	0	0	0	0	0	1	0	1	2	0	13	17	1 (2)

Table 2 (continued)

Gene name	locus Id	Gene size (kb)	Types of variants (SNP + InDel)											% alternate allele (CA allele is taken as reference allele)			Putative FNP (Total)
			5'UTR	3'UTR	Splice variant	Intron	Synonymous	5'UTR premature start codon	Missense & splice variant	Missense	Stop gained	InDel	Total	CA (112) (%)	CF (99) (%)	CC (115) (%)	
PSY	04g22120	6.34	3	19	4	122	18	0	0	9	0	11	186	0	5–92	7–85	9 (9)
	02g19600	2.98	0	0	3	52	14	0	0	10	0	9	88	1–4	1–73	8–48	7 (11)
CRTISO	05g00490	1.5	0	0	0	0	7	0	0	8	0	0	15	24	3–45	1	3 (8)
	10g20040	1.67	0	0	0	0	1	0	0	2	0	0	3	0	0	0	0
CHXB	03g19240	2.28	0	7	3	18	7	0	0	6	0	4	45	0	23	97	1 (2)
	06g24490	2.45	9	8	4	25	5	1	0	8	0	10	70	2	2–89	2–100	4 (9)
NXS	02g04600	3.93	8	0	3	40	6	1	6	6	0	6	76	1	4–100	4–100	4 (7)
CYP450 reductase	04g13960	7.09	0	0	5	301	9	0	1	10	1	25	352	0	1–11	16–96	7 (11)

Chalcone synthase—CHS; *Chalcone isomerase—CHI*; *Dihydrofolate reductase—DFR*; *leucoanthocyanidin reductase—LAR*; *anthocyanidin reductase—ANR*; *UDP glucose: flavonoid-3-O-glucosyltransferase—UFGT*; *Anthocyanin O-methyltransferase—AOMT*; *Phenylalanine ammonia-lyase—PAL*; *4-Coumaric acid: coenzyme A ligase—4CL*; *Hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase—HCT*; *β-ketoacyl-acyl carrier protein synthase- KAS*; *fatty acid translocase—FAT*; *4-coumarate 3-hydroxylase—C3H*; *Caffeoyl-CoA O-methyltransferase 1—CCoAMT*; *Cinnamoyl CoA reductase—CCR*; *putative aminotransferase—pAMT*; *capsaicinoid synthase—CS*; *Phytoene synthase—PSY*; *carotenoid isomerase—CRTISO*; *β-ring hydroxylase—CHXB*; *neoxanthin synthase—NXS*; *Cytochrome P450 reductase—CYP450 reductase*. *Capsicum annuum*—CA; *Capsicum frutescens*—CF; *Capsicum chinense*—CC. The number in brackets indicate number of genotypes

the identification of key biosynthetic pathways which led to the identification of underlying genes using standard ontology-based methods. This would not have been feasible using a targeted metabolite approach. Availability of pangenome (Liu et al. 2023) allowed us an opportunity for allele mining in a set of 326 diverse genotypes. Allele mining for 21 key genes involved in flavonoid, capsaicinoid and carotenoid pathways revealed significant allelic variation for these genes within and between species. For most of the biosynthetic genes, alternate alleles exist. The alternate allele identified for CC and CF can, however, be targeted for designing species-specific allelic markers for at least five of these genes (cut-off > 80%). These markers when used in breeding programs can contribute to desired alleles derived from targeted species. Additionally, for several genes, there is allelic variation between CC and CF. Parallelly, markers targeting stop codons identified in five genes (including *pAMT*) across the three pathways, can be one of the initial steps for testing and if successful, implementing in chilli breeding programs for improvement of fruit traits.

In the case of tomato, a cracking fruit phenotype with lower levels of plant sterols belonging to the brassinosteroid pathway is reported (Rahim et al. 2018). Lower levels of 3β-hydroxysteroid dehydrogenase/decarboxylase isoform2 and δ(24) sterol reductase in this mutant were correlated to cracking fruit type. The MEP/DOXP pathway leading to carotenoid (rhodopin and rhodovibrin identified as key metabolites showing differential expression) and xanthophyll (phoenicoxanthin, adonixanthin, etc.) synthesis was identified as one of the key pathways for which differentially accumulated metabolites were identified. 2-ketospirilloxanthin and 2,2'-diketospirilloxanthin are two metabolites belonging to the spirilloxanthin pathway (Chi et al. 2015). The other two compounds in the pathway which were enriched are rhodovibrin and adonixanthin. While rhodovibin is colourless; adonixanthin is one of the bottlenecks of the maximum accumulation of astaxanthin (Bai et al. 2017). It is known that anthocyanin accumulation is JA inducible (Wasternack and Hause 2013). Tryptophan metabolism has a role in cell enlargement and L-tryptophan was threefold higher in J-41(B) cf AL1. On the other hand, precursors of the brassinosteroid (BR) signalling pathway were at least fivefold higher in KA cf J41(B). It is known that BR has specific effects on differentiation (Müssig 2005); while tryptophan is the precursor of auxin, a hormone well known for influencing cell division, cell elongation and differentiation (Ljung 2013). Shikimate pathway with L-phenylalanine leading to flavones in the flavonoid pathway was identified as another key pathway which showed differential accumulation of metabolites across species. In plants, the biosynthesis of phenolic compounds through aromatic amino acids starts with the shikimate/arogenate pathway (Weaver and Herrmann 1997). The importance of this pathway is demonstrated by the fact that, under normal growth conditions, 20% of the carbon fixed by plants flows through it (Haslam 1993). The formation of aromatic amino acids, phenylalanine, tyrosine and tryptophan, occurs in this pathway. In *C. annuum* seedlings, induction of two enzymes, i.e. shikimate dehydrogenase and peroxidase, is reported in response to copper stress (Díaz et al. 2001). Avocados coated with phenylalanine elicitors are reported to have a better flavour than the control fruit (Saidi et al. 2001). In mango, treatment with phenylalanine enhanced the chilling tolerance of fruit through the regulation of metabolic and defense-related pathways while maintaining high levels of flavonoids, antioxidants, enzyme activity, and volatile aldehydes (Patel et al. 2023). Several reports on targeted metabolite identification in chilli are available and the untargeted approach has only been attempted in *C. annuum* species. In our study we did find certain major metabolites belonging to the four major pathways like colour, flavour and capsaicin development along with fatty acid pathway including glycerophospholipids and sphingolipids which overlap with the published dataset (Das et al. 2021; Lahbib et al. 2021; Cervantes-Hernández et al. 2022), but the relative metabolite comparison across the four heirloom chilli landraces of commercial value and its importance from chilli diversity perspective gives an insight into these complex traits, which has not been reported till date. Identification of key metabolites for colour and flavour, in addition to allelic variation previously reported (Yumnam et al. 2012; Colney et al. 2018) and identified in the current study suggests that polymorphism underlying the biosynthetic genes can be used as biomarkers for further characterisation and utilization of these heirloom landraces for improvement of fruit flavour and colour in chilli. Pangenome revealed that only two genes involved in capsaicin biosynthesis overlapping between *C. annuum* var. *annuum* and *C. baccatum* var. *pendulum*, suggesting that various pungency levels in *C. annuum* var. *annuum* and *C. baccatum* var. *pendulum* were mainly achieved through the selection of different genes in the capsaicin biosynthetic pathway during their independent domestications (Liu et al. 2023). Our untargeted metabolite data allowed us to use the pangenome to identify several other genes including those belonging to capsaicin biosynthetic pathway, which can be targeted for better understanding and improving fruit quality traits. To the best of our knowledge, this study lays the foundation for doing systemic studies to generate the next generation high-quality breeding material in chilli.

Conclusion

Northeast India serves as a secondary center of diversity for chilli peppers. In summary, we examined four 'heirloom' pungent landraces, each differing in fruit colour, shape/size, and texture, and with unique flavours. Using an untargeted approach, we identified a total of 2702 putative metabolites, with 1704 showing differential accumulation.

Key metabolites were identified, along with the biochemical pathways which show differential metabolite accumulation in these chilli landraces. Significant differences in the accumulation of metabolites contributing to membrane phospholipids, including glycerophospholipids and sphingolipids, were identified. Key metabolites and underlying alleles across twenty-one genes and three pathways (flavonoid, capsaicinoid and carotenoid) were identified. The alternate allele identified for CC and CF can be targeted for the design of species-specific allelic markers in at least five of these genes. These markers when used in breeding programs can contribute to desired alleles derived from targeted species. Parallelly, markers targeting stop codons identified in five genes (including *pAMT*) across the three pathways, can be one of the initial steps for testing and if successful, implementing in chilli breeding programs for improvement of fruit traits. This will contribute towards breeding programs in aiding the selection of fruits of the desired trait.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-025-01576-4.

Acknowledgements The research work was funded by CPGSAS, CAU (Imphal), Umiam. Financial support to ALM was provided by CAU (Imphal).

Author contributions Conceptualization, WT; data curation, KY, HS and WT; investigation, ALM, KY, HS, NTM, MR and WT; writing—original draft, ALM, KY and WT; writing—review & editing, KY, NTM, M.R. and WT. All authors have read this version of the manuscript.

Data availability Data is contained within the article and supplementary materials.

Declarations

Competing interests 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.

References

Akhter MJ, Akhter S, Islam S, Sarker MS, Hasan SK (2024) Varietal influence on bioactive compounds and antioxidant activity in chilies during development stages. Heliyon 10(17):e37406

Apeda. Apeda Agriexchange. Web (2021) https://agriexchange.apeda.gov.in/. Accessed 21 Nov 2023

Avuthu T, Sanivarapu H, Prasad K, Sharma N, Sudini HK, Yogendra K (2024) Comparative metabolomics analysis reveals secondary cell wall thickening as a barrier to resist *Aspergillus flavus* infection in groundnut. Physiol Plant 176:e14169

Bai C, Berman J, Farre G, Capell T, Sandmann G, Christou P, Zhu C (2017) Reconstruction of the astaxanthin biosynthesis pathway in rice endosperm reveals a metabolic bottleneck at the level of endogenous β-carotene hydroxylase activity. Transgenic Res 26:13–23

Cervantes-Hernández F, Ochoa-Alejo N, Martínez O, Ordaz-Ortiz JJ (2022) Metabolomic analysis identifies differences between wild and domesticated chili pepper fruits during development (*Capsicum annuum* L.). Front Plant Sci 13:893055

Chamikara MDM, Dissanayake DRRP, Ishan M, Sooriyapathirana SDSS (2016) Dietary, anticancer and medicinal properties of the phytochemicals in chilli pepper (*Capsicum* spp.). Ceylon J Sci 45:5–20

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13(8):1194–1202

Chi SC, Mothersole DJ, Dilbeck P, Niedzwiedzki DM, Zhang H, Qian P, Vasilev C, Grayson KJ, Jackson PJ, Martin EC, Li Y (2015) Assembly of functional photosystem complexes in *Rhodobacter sphaeroides* incorporating carotenoids from the spirilloxanthin pathway. Biochim Biophys Acta-Bioenerg 1847:189–201

Colney L, Tyagi W, Rai M (2018) Morphological and molecular characterization of two distinct chilli cultivars from North Eastern India with special reference to pungency related genes. Sci Hortic 240:1–10

Darko E, Hamow KA, Marček T, Dernovics M, Ahres M, Galiba G (2022) Modulated light dependence of growth, flowering, and the accumulation of secondary metabolites in chilli. Front Plant Sci 13:801656

Das S, Sarkar S, Das M, Banik P, Bhattacharya SS (2021) Influence of soil quality factors on capsaicin biosynthesis, pungency, yield, and produce quality of chili: an insight on *Csy1*, *Pun1*, and *Pun12* signaling responses. Plant Physiol Biochem 166:427–436

Díaz J, Bernal A, Pomar F, Merino F (2001) Induction of shikimate dehydrogenase and peroxidase in pepper (*Capsicum annuum* L.) seedlings in response to copper stress and its relation to lignification. Plant Sci 161:179–188

Du H, Yang J, Chen B, Zhang X, Zhang J, Yang K, Geng S, Wen C (2019) Target sequencing reveals genetic diversity, population structure, core-SNP markers, and fruit shape-associated loci in pepper varieties. BMC Plant Biol 19:1–6

Dutta SK, Singh SB, Vanlalhmangaiha BA, Akoijam RS, Boopathi T, Saha S, Roy S (2018) *Capsicum frutescens* L. landraces of North-East India: from phenotypic diversity perspective of unexplored collection. Proc Natl Acad Sci India B Biol Sci 88:1135–1146

Gaur R, Sharma V, Chhapekar SS, Das J, Kumar A, Yadava SK, Nitin M, Brahma V, Abraham SK, Ramchiary N (2016) Comparative analysis of fruit metabolites and pungency candidate genes expression between bhut jolokia and other *Capsicum* species. PLoS ONE 11:e0167791

Gazette of India: extraordinary (2015) Regarding notification on the 10 crop species (not being extant varieties and farmers' varieties) jasmine, tuberose, papaya, China aster, peach, Japanese plum, strawberry, chilli, bell pepper and paprika, finger millet, foxtail millet for the purpose of registration of varieties. Gazette Notification S.O. 1796(E). Ministry of Agriculture (Department of Agriculture and Cooperation), India. http://plantauthority.gov.in/pdf/chillbellpeppepaprika

Geleta LF, Labuschagne MT, Viljoen CD (2005) Genetic variability in pepper (*Capsicum annuum* L.) estimated by morphological data and amplified fragment length polymorphism markers. Biodivers Conserv 14:2361–2375

González-Pérez S, Garcés-Claver A, Mallor C, Saenz de Miera LE, Fayos O, Pomar F, Merino F, Silvar C (2014) New insights into *Capsicum* spp relatedness and the diversification process of *Capsicum annuum* in Spain. PLoS ONE 9:e116276

Haslam E (1993) Shikimic acid: metabolism and metabolites. Wiley, Chichester

Islam MA, Sharma SS, Sinha P, Negi MS, Neog B, Tripathi SB (2015) Variability in capsaicinoid content in different landraces of *Capsicum* cultivated in north-eastern India. Sci Hortic 183:66–71

Jeong H-J, Jo YD, Park S-W, Kang B-C (2010) Identification of Capsicum species using SNP markers based on high resolution melting analysis. Genome 53:1029–1040

Jha TB, Saha PS, Nath S, Das A, Jha S (2017) Morphological and cytogenetical characterization of 'Dalle Khursani': a polyploid cultivated *Capsicum* of India. Sci Hortic 215:80–90

Lahbib K, Dabbou S, Bnejdi F, Pandino G, Lombardo S, El Gazzah M, El Bok S (2021) Agro-morphological, biochemical and antioxidant characterization of a Tunisian chili pepper germplasm collection. Agriculture 11:1236

Lee HY, Ro NY, Patil A, Lee JH, Kwon JK, Kang BC (2020) Uncovering candidate genes controlling major fruit-related traits in pepper via genotype-by-sequencing based QTL mapping and genome-wide association study. Front Plant Sci 11:1–15

Liu F, Zhao J, Sun H, Xiong C, Sun X, Wang X, Wang Z, Jarret R, Wang J, Tang B, Xu H (2023) Genomes of cultivated and wild *Capsicum* species provide insights into pepper domestication and population differentiation. Nat Commun 14:5487

Ljung K (2013) Auxin metabolism and homeostasis during plant development. Development 140:943–950

López Castilla LD, Garruña Hernández R, Castillo Aguilar CD, Martínez-Hernández A, Ortiz-García MM, Andueza-Noh RH (2019) Structure and genetic diversity of nine important landraces of Capsicum species cultivated in the Yucatan Peninsula, Mexico. Agronomy 9:376

McCoy J, Martínez-Ainsworth N, Bernau V, Scheppler H, Hedblom G, Adhikari A, McCormick A, Kantar M, McHale L, Jardón-Barbolla L, Mercer KL (2023) Population structure in diverse pepper (*Capsicum* spp.) accessions. BMC Res Notes 16:20

Müssig C (2005) Brassinosteroid-promoted growth. Plant Biol 7(2):110–117

Naseem M, Rana RM, ul Hassan M, Khan MA, Binobead MA, Elsadek MF, Elsalahy HH, Iqbal R (2024) High performance liquid chromatography (HPLC) based genetic diversity profiling of chilli germplasm for fruit pungency and phytochemical contents. BMC Plant Biol 24:368

Naves ER, Scossa F, Araújo WL, Nunes-Nesi A, Fernie AR, Zsögön A (2022) Heterosis for capsacinoids accumulation in chili pepper hybrids is dependent on parent-of-origin effect. Sci Rep 12:14450

Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques PÉ, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396

Paredes Andrade NJ, Monteros-Altamirano A, Tapia Bastidas CG, Sørensen M (2020) Morphological, sensorial and chemical characterization of chilli peppers (*Capsicum* spp.) from the CATIE genebank. Agronomy 10:1732

Patel MK, Fanyuk M, Feyngenberg O, Maurer D, Sela N, Ovadia R, Oren-Shamir M, Alkan N (2023) Phenylalanine induces mango fruit resistance against chilling injuries during storage at suboptimal temperature. Food Chem 405:134909

Peñuela M, Arias LL, Viáfara-Vega R, Rivera Franco N, Cárdenas H (2021) Morphological and molecular description of three commercial Capsicum varieties: a look at the correlation of traits and genetic distancing. Genet Resour Crop Evol 68:261–277

Perry L, Dickau R, Zarrillo S, Holst I, Pearsall DM, Piperno DR, Berman MJ, Cooke RG, Rademaker K, Ranere AJ, Raymond JS (2007) Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. Science 315:986–988

Pickersgill B, Heiser CB, McNeill J (1979) Numerical taxonomic studies on variation and domestication in some species of *Capsicum*. In: Hawkes JG, Lester RN, Skelding AD (eds) The biology and taxonomy of Solanaceae. Linnean Society Symposium 7. Academic, London, pp 678–700

Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinform 11:1–11

Rahim MA, Jung HJ, Afrin KS, Lee JH, Nou IS (2018) Comparative transcriptome analysis provides insights into dwarfism in cherry tomato (*Solanum lycopersicum* var. cerasiforme). PLoS ONE 13:e0208770

Reddy UK, Almeida A, Abburi VL, Alaparthi SB, Unselt D, Hankins G, Park M, Choi D, Nimmakayala P (2014) Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in *Capsicum annuum* L. collections. PLoS ONE 9:e86393

Rivera A, Monteagudo AB, Igartua E, Taboada A, García-Ulloa A, Pomar F, Riveiro-Leira M, Silvar C (2016) Assessing genetic and phenotypic diversity in pepper (*Capsicum annuum* L.) landraces from North-West Spain. Sci Hortic 203:1–1

Saidi L, Duanis-Assaf D, Galsarker O, Maurer D, Alkan N, Poverenov E (2001) Elicitation of fruit defense response by active edible coatings embedded with phenylalanine to improve quality and storability of avocado fruit. Postharvest Biol Technol 174:111442

Sanatombi K, Sharma GJ (2008) Capsaicin content and pungency of different *Capsicum* spp. cultivars. Not Bot Horti Agrobot 36:89–90

Summer LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics 3:211–221

von Steimker J, Tripodi P, Wendenburg R, Tringovska I, Nankar AN, Stoeva V, Pasev G, Klemmer A, Todorova V, Bulut M, Tikunov Y (2024) The genetic architecture of the pepper metabolome and the biosynthesis of its signature capsianoside metabolites. Curr Biol 34:4209–4223

Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG (2011) Metabolite biodiversity in pepper (*Capsicum*) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. Phytochemistry 72:1358–1370

Wahyuni Y, Ballester AR, Tikunov Y, de Vos RC, Pelgrom KT, Maharijaya A, Sudarmonowati E, Bino RJ, Bovy AG (2013a) Metabolomics and molecular marker analysis to explore pepper (*Capsicum* sp.) biodiversity. Metabolomics 9:130–144

Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG (2013b) Secondary metabolites of *Capsicum* species and their importance in the human diet. J Nat Prod 76:783–793

Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. Ann Bot 111:1021–1058

Weaver LM, Herrmann KM (1997) Dynamics of the shikimate pathway in plants. Trends Plant Sci 2:346–351

Yumnam JS, Tyagi W, Pandey A, Meetei NT, Rai M (2012) Evaluation of genetic diversity of chilli landraces from North Eastern India based on morphology, SSR markers and the *Pun1* locus. Plant Mol Biol Rep 30:1470–1479