Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2025), pp. 1–13 doi: 10.1111/pbi.70171

High-level production of health-beneficial glucoraphanin by multiplex editing of *AOP2* gene family in mustard

Pravin Kumar and Naveen C. Bisht*

National Institute of Plant Genome Research, New Delhi, India

Received 4 October 2024;
revised 30 April 2025;
accepted 20 May 2025.
**Correspondence* (Tel +91-11-26735183; fax +91-11-26741658; email ncbisht@nipgr.ac.in)

Keywords: *AOP2*, genome editing, glucoraphanin, glucosinolates, *Brassica juncea*, sulforaphane.

Summary

Intake of glucosinolates through the consumption of cruciferous vegetables has been associated with numerous health benefits. In recent decades, glucosinolate glucoraphanin has gained a lot of attention, as its hydrolysis product (sulforaphane) is known to possess numerous health-promoting benefits, including anti-cancer and chemopreventive activities. However, due to the low availability of glucoraphanin in most of the cultivated *Brassica* crops (except broccoli), there is an increasing interest in many laboratories around the world to manipulate the glucosinolate profile for human benefit. Here, we report the high-level production of health-beneficial glucoraphanin by CRISPR/Cas9 editing of the *ALKENYL HYDROXALKYL PRODUCING 2* (*BjuAOP2*) gene family, displaying distinct expression profiles in the allotetraploid mustard, *Brassica juncea*. Multiplex editing of five *BjuAOP2* homologues, using four gRNAs, provided glucoraphanin accumulation up to 41.60, 75.10, 59.21 and 27.64 μmoles/g dry weight in sprouts, microgreens, seeds and leaves, respectively, of the transgene-free *BjuAOP2*-edited lines, while providing a significant reduction of the anti-nutritional and goitrogenic alkenyl glucosinolates including progoitrin. The glucoraphanin enhancement in *BjuAOP2*-edited lines was found to be dose-dependent, wherein loss-of-function mutations in *BjuAOP2.A09* and *BjuAOP2.B01* homologues had a more prominent effect. The transgene-free *BjuAOP2*-edited lines were stable for high glucoraphanin and performed at par with the wild-type plants for various seed quality and yield parameters when tested under containment conditions in the field. The development of high-glucoraphanin mustard will help its adoption as a global superfood with health-promoting benefits and as a bioactive source of high-value sulforaphane for industrial production.

Introduction

The crops belonging to the Brassicaceae (or cruciferous) family are cultivated globally as oilseed, vegetables, condiments and biofumigants. These crops are blessed with the presence of a class of lineage-specific nitrogen and sulfur-containing specialized metabolites called glucosinolates (Halkier and Gershenzon, 2006). Glucosinolates are derived from amino acids, with more than 130 structures known to date (Blažević *et al.*, 2020). Glucosinolates upon hydrolysis by myrosinases (thioglucosidase, EC 3.2.1.147) produce several bioactive compounds such as isothiocyanates, thiocyanates and nitriles, which possess various biological activities important for plant defence and animal health (Fahey *et al.*, 2012; Hopkins *et al.*, 2009).

Glucoraphanin (4-methylsulfinylbutyl glucosinolate, 4MSOB) is one of the widely studied glucosinolates that possess several health benefits. Sulforaphane (4MSOB-ITC), an isothiocyanate of glucoraphanin, has gained a lot of attention in recent decades due to its various beneficial pharmacological properties including anti-carcinogenic, anti-oxidant, anti-inflammatory, anti-apoptotic, anti-microbial, anti-viral and anti-aging activities investigated in *in vitro*, *in vivo* and clinical studies (Baralić *et al.*, 2024; Clarke *et al.*, 2008; Fahey *et al.*, 1997; Houghton, 2019; Zhang *et al.*, 1992, 2024). The anti-cancer and chemopreventive activities of sulforaphane are attributed to its ability to activate the human Nrf2 (nuclear erythroid 2-related factor-2) signalling pathways, effective inhibition of cell cycle progression, deactivation of carcinogen-activating phase-I enzymes, activation of carcinogen-detoxifying phase-II enzymes, free radical scavenging, histone deacetylation and topoisomerases suppression (reviewed by Clarke *et al.*, 2008; Higdon *et al.*, 2007; Kaiser *et al.*, 2021; Liu *et al.*, 2024; Russo *et al.*, 2018). Besides, sulforaphane is an effective bioactive compound as it is readily available in the blood (80% bioavailability) due to its low molecular weight compared to many other plant-derived metabolites (Hanlon *et al.*, 2008; Houghton, 2019).

However, despite having several health-beneficial properties and higher bioavailability, glucoraphanin is not consumed in sufficient amounts that can alter the health biomarkers and suffice the recommended dietary requirements (Kaiser *et al.*, 2021; Yagishita *et al.*, 2019). Among the cruciferous vegetables, broccoli (*Brassica oleracea* var. *italica*) accumulates high amounts of glucoraphanin, whereas other members such as Chinese kale, cabbage and brussels sprout possess comparably lower glucoraphanin (Kushad *et al.*, 1999; Verkerk *et al.*, 2009). A significant leap was made with the development of the broccoli hybrid Beneforte®, commercialized as a 'superfood', which accumulates

Please cite this article as: Kumar, P. and Bisht, N.C. (2025) High-level production of health-beneficial glucoraphanin by multiplex editing of *AOP2* gene family in mustard. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.70171.

glucoraphanin in the range of 21.7–30.1 μmoles/g dry weight (hereafter DW), the highest known to date (Sivapalan *et al*., 2018; Traka *et al*., 2013). As broccoli cultivation is limited to temperate climatic conditions, there is an urgent need for engineering the high-value glucoraphanin in other globally cultivated *Brassica* crops for human consumption.

Brassica juncea (Indian mustard) is an emerging oilseed crop, which is also consumed globally as a leafy vegetable and salad (sprouts and mustard microgreens). In recent decades, the crop has been a good alternative to rapeseed (Canola) due to its better adaptability to wider climatic conditions, higher abiotic and biotic tolerance, and pod-shattering tolerance (Wang *et al*., 2007; Zandberg *et al*., 2022). However, *B. juncea* contains very high amounts of glucosinolates that are considered anti-nutritional and this reduces the oil and oil cake value in the global market (Augustine *et al*., 2013; Cools and Terry, 2018; Mann *et al*., 2023). We recently profiled the core collection of *B. juncea* diversity panel, comprising 233 accessions collected worldwide, and reported that mustard germplasm contains very high amounts of undesirable goitrogenic alkenyl glucosinolates (gluconapin and sinigrin) whereas the health-beneficial glucoraphanin is present in traces, despite possessing functional aliphatic glucosinolate biosynthesis machinery (Augustine and Bisht, 2017; Gohain *et al*., 2021).

Briefly, aliphatic glucosinolate biosynthesis from amino acid methionine occurs through three major steps, namely, side chain elongation, core structure formation and side chain modifications (Figure S1). In the model plant *Arabidopsis thaliana*, 2-oxoglutarate-dependent dioxygenases encoding genes viz., *AOP1* (*At4g03070*), *AOP2* (*At4g03060*) and *AOP3* (*At4g03050*), involved in side chain modification step, are known to shape the natural variations for methylsulfinylalkyl, alkenyl and hydroxyalkyl glucosinolates (Giamoustaris and Mithen, 1996; Kliebenstein *et al*., 2001). *AOP1* is believed to be an ancestral gene, which through a series of gene duplication events gave rise to *AOP2* and *AOP3* (Gao *et al*., 2004). While the function of *Arabidopsis* AOP1 is not yet characterized, AOP2 and AOP3 catalyze the formation of alkenyl and hydroxyalkyl glucosinolates, respectively (Kliebenstein *et al*., 2001).

The crops belonging to the Brassicaceae family including *B. juncea* and its progenitor species *Brassica rapa* and *Brassica nigra* possess multiple copies of functional *AOP2* genes, leading to a high alkenyl glucosinolate pool (Augustine and Bisht, 2015; Chen *et al*., 2023; Gao *et al*., 2004). In contrast, broccoli possesses non-functional *AOP2* genes, due to a loss of the conserved 2OG-FeII_Oxy domain, which explains the high accumulation of health-promoting glucoraphanin in broccoli (Cai *et al*., 2024). Since *B. juncea* contains high amounts of anti-nutritional goitrogenic alkenyl glucosinolates, we hypothesized that *AOP2* genes are catalytically efficient and can be targeted for engineering high amounts of health-beneficial glucoraphanin for human consumption.

Here, we report the high-level production of chemopreventive glucoraphanin in *B. juncea* through CRISPR/Cas9-based multiplex editing of five *AOP2* homologues. We also performed the screening and inheritance of mutations across generations, the development of transgene-free lines with significantly high glucoraphanin content in seeds, leaves, sprouts and microgreens, and their agronomical assessment under contained growth conditions. The *B. juncea* crop biofortified with up to 75.10 μmoles/g DW glucoraphanin content is the highest known to date in any of the Brassicaceae crops and will enhance its adoption as a leafy green vegetable crop and as a bioactive source of high-value sulforaphane for industrial production.

Results

Brassica juncea AOP2 homologues share a close evolutionary relationship but distinct expression patterns

A high-yielding, high-glucosinolate containing *B. juncea* variety Varuna, with negligible glucoraphanin (4MSOB) content in leaves and seeds, was selected for the CRISPR/Cas9-based editing of the *B. juncea AOP2* (*BjuAOP2*) homologues. Using the *Arabidopsis AOP2* gene as a query sequence, genomic sequences of five *BjuAOP2* homologues, namely *BjuAOP2.A02*, *BjuAOP2.A03*, *BjuAOP2.A09*, *BjuAOP2.B01* and *BjuAOP2.B03*, were retrieved using *B. juncea* cv. Varuna genome assembly (Paritosh *et al*., 2021, 2024). The nomenclature of the *BjuAOP2* genes was adopted as per the *Brassica* gene nomenclature, wherein A and B denote the two subgenomes of the allotetraploid *B. juncea* and their respective chromosomal locations in the *B. juncea* cv. Varuna assembly (Østergaard and King, 2008; Paritosh *et al*., 2021, 2024).

The full-length genomic and coding DNA sequence (CDS) of all *BjuAOP2* homologues ranged from 1781 to 5608 bp and 1299 to 1499 bp, respectively (Table S1). Analysis of the genomic structure of *BjuAOP2* homologues revealed that all *BjuAOP2* homologues have retained three exons and two introns while showing a higher variation in the length of introns (Table S1). Notably, the *BjuAOP2.A02* exhibited an insertion of 213 bp in its second exon, leading to an aberrant termination of the encoded protein (Figure 2a). The insertion in *BjuAOP2.A02* is unique to *B. juncea* cv Varuna, contrary to its counterpart in the progenitor *B. rapa* genome (Figure S2). Sequence alignment of CDS revealed that *BjuAOP2* homologues shared 67.4%–69.4% identity with the *AtAOP2* while sharing a higher level of CDS identity (76.1%–91.7%) among themselves (Figure S3).

Further to establish the evolutionary relationship among *BjuAOP2* homologues, a phylogenetic tree was constructed using the encoded amino acid sequences of *B. juncea* and *Arabidopsis AOP* genes. In a maximum likelihood tree, all five BjuAOP2 proteins were grouped with the *Arabidopsis* AtAOP2, while AtAOP1 was an outgroup protein (Figure 1a). In general, the 'A' and 'B' subgenome AOP2 proteins of *B. juncea* were grouped along with their *B. rapa* (Bra) and *B. nigra* (BniB) orthologs, respectively, supported by high bootstrap values, suggesting their close evolutionary relationship. Notably, (BjuAOP2.A09 and BjuAOP2.B03) and (BjuAOP2.A02 and BjuAOP2.B01) formed two orthologous groups, while BjuAOP2.A03 formed an independent clade.

The presence of multiple *BjuAOP2* homologues promoted us to determine their tissue-specific expression patterns using real-time quantitative PCR (RT-qPCR) analysis. The five *BjuAOP2* homologues have retained distinct expression patterns in different tissue types of *B. juncea* (Figure 1b). In general, the two B subgenome-specific homologues have a higher transcript abundance compared to the A subgenome-specific homologues. This somewhat correlates with the quantitative abundance of the various cis-acting regulatory elements present in their 2 kb promoter regions, as identified using the PlantCARE database (Figure S4; Lescot *et al*., 2002). While *BjuAOP2.B01* and *BjuAOP2.B03* displayed similar expression patterns, *BjuAOP2.B03* was the highly expressed gene across all the tested tissue types.

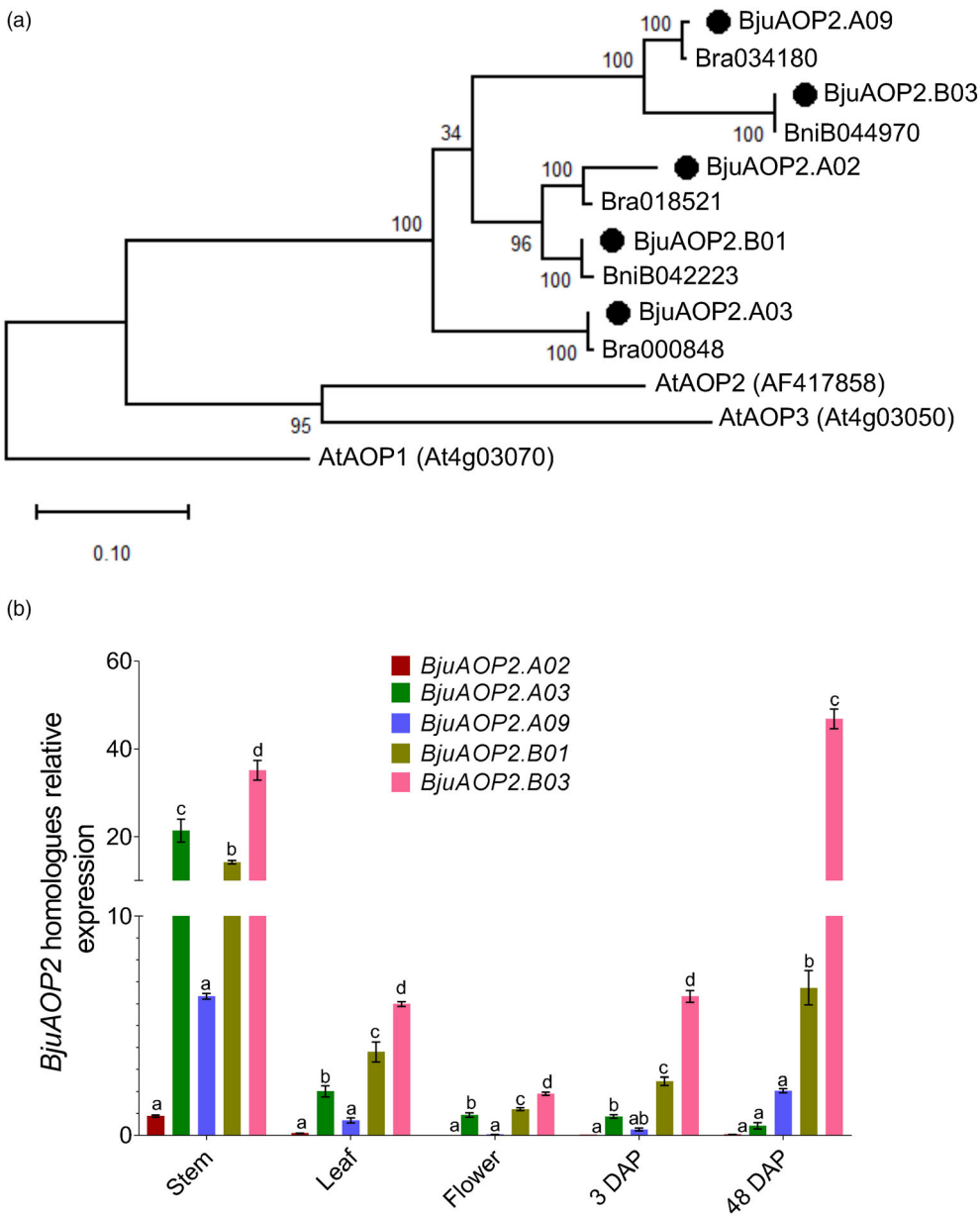

Figure 1 Molecular and expression analysis of *BjuAOP* homologues. (a) Evolutionary relationship of the AOP2 proteins of *B. juncea* (BjuAOP2, marked in dark circles), progenitors *B. rapa* (BraXXXXXX) and *B. nigra* (BniBXXXXXX) with the *Arabidopsis* AOP proteins. Phylogenetic analysis was inferred using the maximum likelihood method in MEGA11 software. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The phylogenetic tree is drawn to scale, with branch lengths measured in the number of substitutions per site. (b) Expression analysis of *BjuAOP2* homologues in different tissue types of *B. juncea*. RT-qPCR experiments were performed using *BjuActin* as the reference control (Chandna *et al.*, 2012), and error bars represent ±SE of the mean (n = 3). Different letters indicate significant expression differences among the *BjuAOP2* homologues within each developmental stage, calculated using one-way analysis of variance ($P < 0.05$) following Tukey's post hoc test.

Among the A subgenome homologues, all three homologues exhibited low transcript abundance in all the tissue types analysed, except stem tissue where *BjuAOP2.A03* had a relatively higher transcript abundance.

Multiplex editing of *BjuAOP2* homologues leads to the generation of high glucoraphanin mustard

To generate genetic lesions in *BjuAOP2* homologues, CRISPR/Cas9-based multiplex editing was performed. A total of four guide RNAs (gRNAs) targeting the first exon of the five *BjuAOP2* homologues was designed using CRISPR-P v2.0. The gRNA1 targets *BjuAOP2.B03*; gRNA2 targets *BjuAOP2.A02*, *BjuAOP2.A03*, *BjuAOP2.A09* and *BjuAOP2.B01*; gRNA3 targets *BjuAOP2.A02*, *BjuAOP2.A09* and *BjuAOP2.B03*; whereas gRNA4 targets *BjuAOP2.A03*, *BjuAOP2.A09* and *BjuAOP2.B03* (Figure 2a; Figure S5, Table S2). All four sgRNAs (ptAtU6-gRNA-scaffold) fragments were amplified as independent expression cassettes (Figure S6) and cloned in a pPZP200 binary vector, containing ptCsVMV-*SpCas9*-pA and 35Sde-*bar*-pA expression cassettes (Mann *et al.*, 2023), to develop the final *BjuAOP2*(GEd) construct (Figure 2b). The construct was used for *Agrobacterium*-mediated genetic transformation of a high

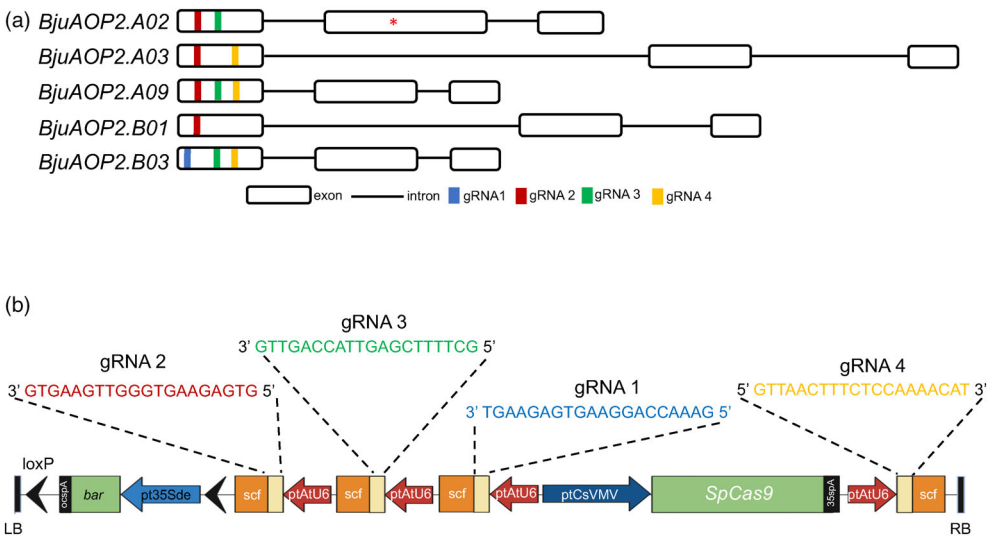

Figure 2 Genomic location of gRNA target sites and generation of *BjuAOP2*(GEd) construct. (a) Genomic location of target sites of four gRNAs viz., gRNA1, gRNA2, gRNA3 and gRNA4 targeting the first exon of the *BjuAOP2* homologues. The gene structure and alignment of *BjuAOP2* homologues are provided in Table S1 and Figures S2 and S5, respectively. An asterisk in exon 2 of *BjuAOP2.A02* indicates premature termination of the encoded protein in *B. juncea* cv. Varuna, due to an insertion of 213 bp sequence. (b) T-DNA map showing SpCas9-based *BjuAOP2*(GEd) transformation construct used for targeting the *BjuAOP2* genes in *B. juncea*. The four sgRNA fragments were cloned sequentially into the multiple cloning sites of pZP200:35Sde:*bar*:*SpCas9* vector, containing *bar* gene as the plant selection marker. A description of gRNAs used for the development of the *BjuAOP2*(GEd) construct is provided in Table S2.

glucosinolate containing *B. juncea* cv. Varuna, following the protocol established in the laboratory (Augustine *et al.*, 2013). Four independent T_0 transgenic events (L1, L3, L5 and L7), having normal growth and good seed set, were selected for further analysis.

The fractional glucosinolate content in T_1 seeds (obtained from T_0 events) was estimated using HPLC. The seeds of *B. juncea* wild-type (WT) cv. Varuna accumulate very high glucosinolates (144.46 ± 6.7 μmoles/g DW), wherein the non-desirable alkenyl glucosinolates, namely, sinigrin (SIN), gluconapin (GNA) and glucobrassicanapin (GBN), constitute the major components (up to 98% of the total pool), whereas methylsulfinylalkyl glucosinolates, namely, 3-methylsulfinylpropyl (3MSOP), 4-methylsulfinylbutyl (4MSOB, glucoraphanin) and 5-methylsulfinylpentyl (5MSOP), are present in trace amounts (Figure 3a, Table S3). In contrast, the T_1 seeds of the *BjuAOP2*(GEd) events accumulated 11.41–71.14 μmoles/g DW of methylsulfinylalkyl glucosinolates. In particular, the content of the health-beneficial glucosinolate, glucoraphanin (4MSOB) in the T_1 seeds ranged from 7.54 to 56.43 μmoles/g DW compared to 0.56 μmoles/g DW present in the wild-type Varuna (Figure 3a; Table S3). The glucoraphanin fraction accounted for 5.68%–62.78% of the total seed glucosinolates in the primary T_0 transformants compared to 0.38% present in the wild-type Varuna. Furthermore, the *BjuAOP2* (GEd) events exhibited a concomitant reduction of the anti-nutritional alkenyl glucosinolates compared to that present in the wild-type Varuna.

Differential enhancement of glucoraphanin content across T_0 events led us to investigate the mutation profile in *BjuAOP2* homologues. Homologue-specific primers were designed to amplify the targeted genomic regions of *BjuAOP2* homologues, and Sanger sequencing chromatograms were used to determine the mutation efficiency, mutation types, allelic status and the zygosity of mutations (Figure 3b). The CRISPR-induced mutations in all four T_0 events included different ranges of deletions (−1 bp to −219 bp) and insertions (+1 bp to +7 bp) across the *BjuAOP2* target sites (Table S4). The use of four sgRNAs allowed us to obtain 75%–100% mutation frequency across the five *BjuAOP2* homologues, causing a variety of chimeric, biallelic homozygous (BHo) and biallelic heterozygous (BHt) mutations at the target sites (Figure 3b; Table S5).

The mutation spectrum in *BjuAOP2* homologues was compared between L1 (containing approx. 62.78% glucoraphanin) and L3, L5 and L7 (with <11% glucoraphanin). L1 has frame-shift mutations in both the alleles of all five *BjuAOP2* homologues in BHo, BHt and chimeric (Chi) conditions, leading to a complete loss-of-gene function. In the other three T_0 events, with low seed glucoraphanin content, a few *BjuAOP2* homologues remained either unedited (WT) or had a shorter deletion in the multiples of 3 bp (~WT), creating alleles with in-frame deletions, which might have retained the native function of the otherwise edited *BjuAOP2* alleles in the T_0 events (Table S4). Overall, our initial data suggested that frameshift mutations in multiple *BjuAOP2* homologues are necessary for obtaining a high-glucoraphanin accumulation in the allotetraploid *B. juncea*.

The high glucoraphanin content is attributed to a dose-dependent effect of *BjuAOP2* mutations

The inheritance of the high-glucoraphanin trait in all four T_0 events was analyzed in the advanced generations. Several transgene-free progenies (without *bar* and *SpCas9* expression cassettes) for each T_0 event were identified in the advanced T_2 generation based on their 'Basta' sensitive phenotype and negative amplification for the '*bar*' gene (Figure S7, Table S6). T_3 seeds from the transgene-free T_2 progenies of L1, L3, L5 and L7 accumulated variable glucoraphanin content ranging from 38.74 to 64.10 μmoles/g DW, 0.21 to 56.91 μmoles/g DW, 0.58 to 51.99 μmoles/g DW and 0.21 to 50.06 μmoles/g DW, respectively, compared to the wild-type Varuna (0.29 μmoles/g

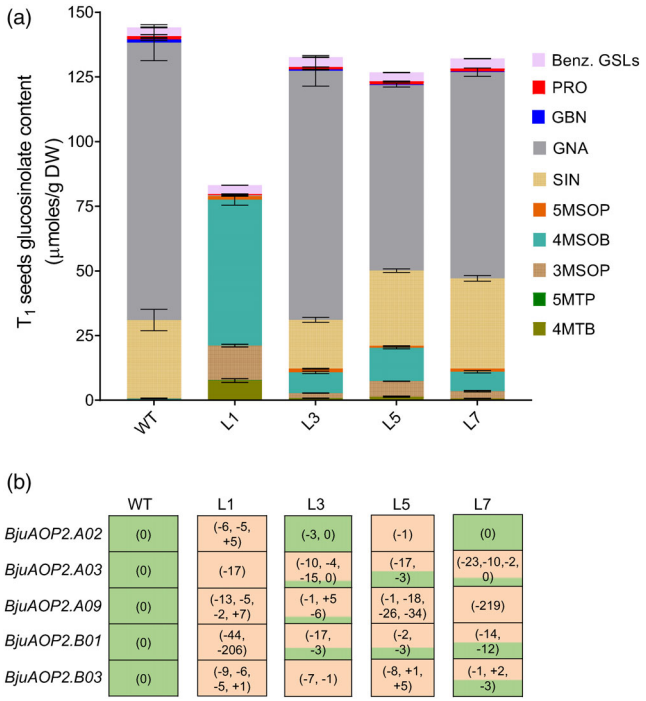

Figure 3 Glucosinolate profiles and mutation analysis of *BjuAOP2*-edited *B. juncea* lines. (a) Comparison of fractions of beneficial seed glucoraphanin with undesirable glucosinolates in T_1 seeds. Bar plot showing the mean of total seeds glucosinolates fractions (µmoles/g DW) in T_1 seeds of four T_0 transgenic events. Each bar represents the mean value ±SE estimated using HPLC ($n > 4$). Total and individual glucosinolates from T_1 seeds are provided in Table S3. (b) Analysis of CRISPR/Cas9-induced *BjuAOP2* editing pattern in four independent T_0 events. Different colour coding shows the functionality (based on mutations) in both the alleles of each *BjuAOP2* homologue. The alleles with frame-shift mutations are marked in orange; alleles which are unedited (0) or having in-frame mutations (short deletion in multiple of 3 bp) are shown in green; and a gradient of orange and green are the cases where the edited alleles have a mix of frame-shift mutations, in-frame mutations or remain unedited. Mutation data for each gRNA target site in T_0 events is provided in Table S4. 3MSOP, 3-methylsulfinylpropyl; 4MSOB, 4-methylsulfinylbutyl (glucoraphanin); 4MTB, 4-methylthiobutyl; 5MSOP, 5-methylsulfinylpentyl; 5MTP, 5-methylthiopentyl; GBN, 4-pentenyl (glucobrassicanapin); GNA, 3-butenyl (gluconapin); PRO, 2-hydroxy-3-butenyl (progoitrin); SIN, 2-propenyl (sinigrin); Benz. GSLs include both indolic and benzenic glucosinolates. WT, wild-type.

DW) (Table S7). At the percentage level, the increased glucoraphanin content in T_3 seeds of L1, L3, L5 and L7 progenies ranged from 65.69% to 71.75%, 0.17% to 63.03%, 0.61% to 74.98% and 0.20% to 71.31%, respectively, compared to 0.23% in Varuna (Table S7). Thus, except for L1, progenies from all other events showed a wide variation of seed glucoraphanin content in advanced generations, which could be attributed to the segregation of mutated *BjuAOP2* homologues.

The wide range of seed glucoraphanin content in the advanced T_2 generation provided us with an opportunity to investigate the segregation of CRISPR/Cas9-induced mutations vis-à-vis the contribution of each *BjuAOP2* homologue for the seed glucoraphanin content. Based on the percentage of seed glucoraphanin accumulation, transgene-free T_2 lines were categorized into four groups namely, Group I (0%–20%), Group II (21%–40%), Group III (41%–60%) and Group IV (>60%) (Figure 4a; Table S7). Subsequently, 17 transgene-free T_2 lines representing all four groups were analysed for the segregation of *BjuAOP2* mutations. The analysis suggested a dose-dependent effect of the *BjuAOP2* mutations on the seed glucoraphanin trait (Figure 4b, Table S8). For example, frameshift mutations in *BjuAOP2.A03*, *BjuAOP2.B01* and *BjuAOP2.B03* alleles (2–5 mutated alleles) could only provide up to 33% glucoraphanin content of the total seed glucosinolates (Group I and II; Figure 4b). Mutations in *BjuAOP2.A09* further enhanced the glucoraphanin content up to around 60% of the total seed glucosinolates (6–7 mutated alleles; Group III). However, loss-of-function mutations in four *BjuAOP2* homologues namely *BjuAOP2.A03*, *BjuAOP2.A09*, *BjuAOP2.B01* and *BjuAOP2.B03* could only generate lines with >70% seed glucoraphanin content (8 mutated alleles; Group IV). As expected, CRISPR-induced mutations in the truncated *BjuAOP2.A02* homologue did not confer any further enhancement to the glucoraphanin content (Figures 1b, 2a and 4b). A representative sequence alignment of mutated *BjuAOP2* alleles in the transgene-free T_2 lines from each segregating group is provided in Figure 4c.

The transgene-free *BjuAOP2*-edited lines display a stable inheritance of high-glucoraphanin with a concomitant reduction of anti-nutritional glucosinolates

In order to check the stability of the high-glucoraphanin trait, a total of 12 transgene-free *BjuAOP2*-edited lines, belonging to the Group III and IV categories, were analyzed for glucoraphanin content in the sprouts, microgreens, leaves and seeds in the advanced T_3 generation plants. The germinating seeds (2 days old sprouts) of the *BjuAOP2*-edited lines accumulated mean glucoraphanin in the range of 21.97–41.60 µmoles/g DW, accounting for approximately 62% of the total glucosinolates in sprouts (Figure 5a, Table S9). Further, 5 days old microgreens of the *BjuAOP2*-edited lines accumulated mean glucoraphanin in the range of 62.91–75.10 µmoles/g DW, which accounted for approximately 76% of the total glucosinolates (Figure 5b,

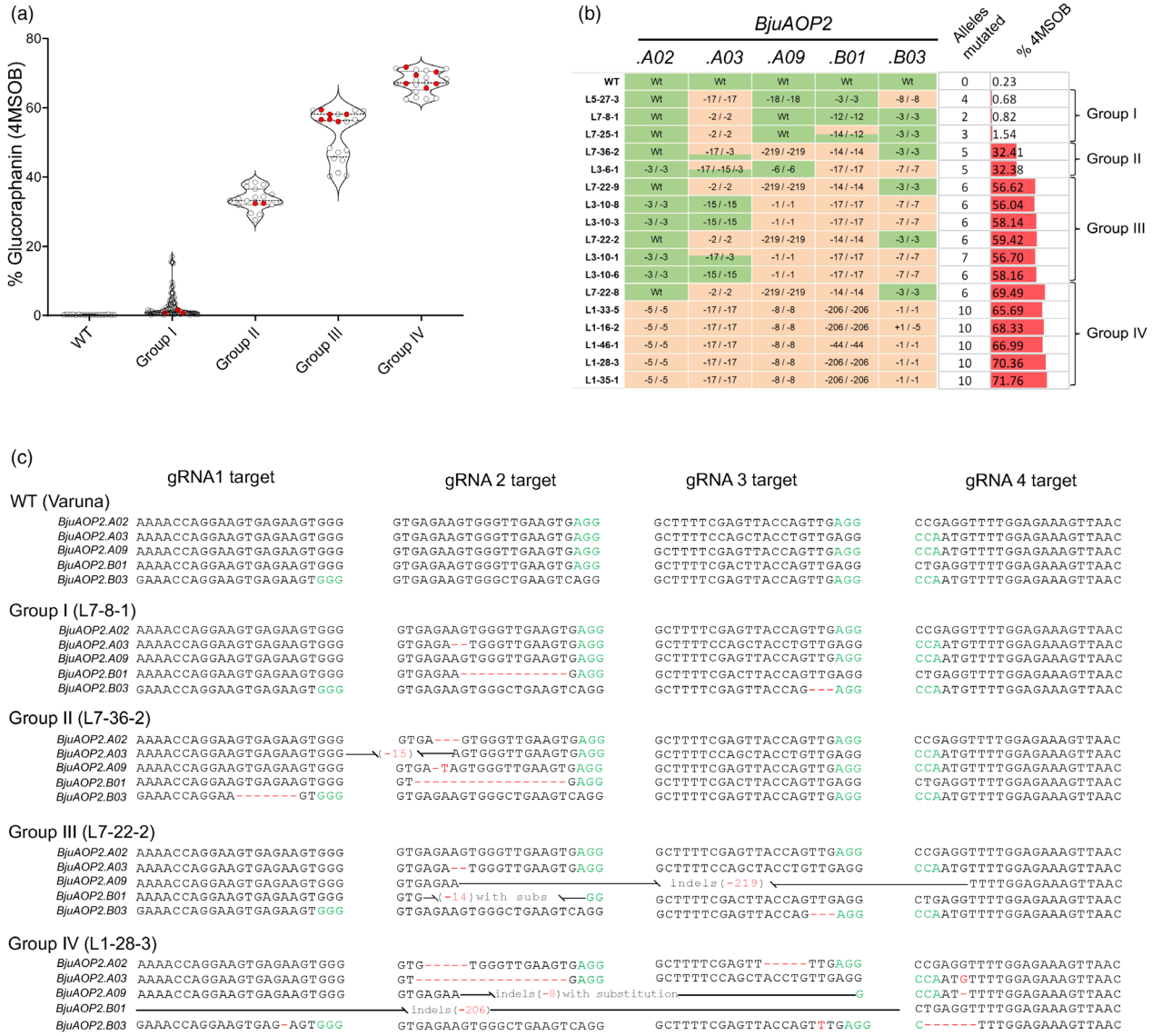

Figure 4 Analysis of mutations in *BjuAOP2* homologues and glucoraphanin accumulation in transgene-free lines. (a) Violin plot showing four distinct groups (Group I–IV) of transgene-free T_2 lines based on percent glucoraphanin pool in T_3 seeds. Each circle represents seeds glucoraphanin content from individual T_2 plants. The coloured circles represent T_2 plants that were genotyped for the identification of mutations. Data on total and individual glucosinolates from T_3 seeds are provided in Table S7. (b) Analysis of editing profiles and genetic zygosity in transgene-free *BjuAOP2*-edited T_2 lines having a range of glucoraphanin (4MSOB) accumulation (% of total seed glucosinolate content). Different colour coding shows the functionality (based on mutations) in both the alleles of each *BjuAOP2* homologue. The alleles with frame-shift mutations are marked in orange; alleles which are unedited (0) or having in-frame mutations (short deletion in multiple of 3 bp) are shown in green; and the gradient of orange and green are the cases where the edited alleles have a mix of frame-shift mutations, in-frame mutations or remain unedited. The mutation profile of transgene-free *BjuAOP2*-edited T_2 lines is provided in Table S8. (c) A representative sequence alignment showing the mutations at the target sites of *BjuAOP2* homologues in transgene-free T_2 lines representing each group.

Table S10). Young leaves of 30–35 days field-grown *BjuAOP2*-edited lines accumulated mean glucoraphanin in the range of 19.34–27.64 µmoles/g DW, which corresponds to approximately 72% of the leaf total glucosinolate content (Figure 5c, Table S11). Further, the transgene-free T_4 seeds from the *BjuAOP2*-edited lines accumulated mean glucoraphanin in the range of 43.13–59.21 µmoles/g DW, that is approximately 65% of the total seed glucosinolates (Figure 5d, Table S12). Thus, editing of *BjuAOP2* homologues provided a significant enhancement of glucoraphanin contents throughout the tested plant organs compared to the wild-type Varuna. It is also worth noting that CRISPR/Cas9-mediated complete knock-out of *BjuAOP2* homologues provided up to twofold higher glucoraphanin content in all the tested tissue types compared to the *BjuAOP2* (RNAi) lines generated in our earlier study (Augustine and Bisht, 2015; Table S12).

We further quantified the precursor and product glucosinolates of AOP2 in the transgene-free *BjuAOP2*-edited lines. The precursor methylthioalkyl glucosinolates, namely, 3-methylthiopropyl (3MTP), 4-methylthiobutyl (4MTB) and 5-methylthiopentyl (5MTP), which were otherwise not detected in the wild-type Varuna, were accumulated in sprouts (4.18–9.06

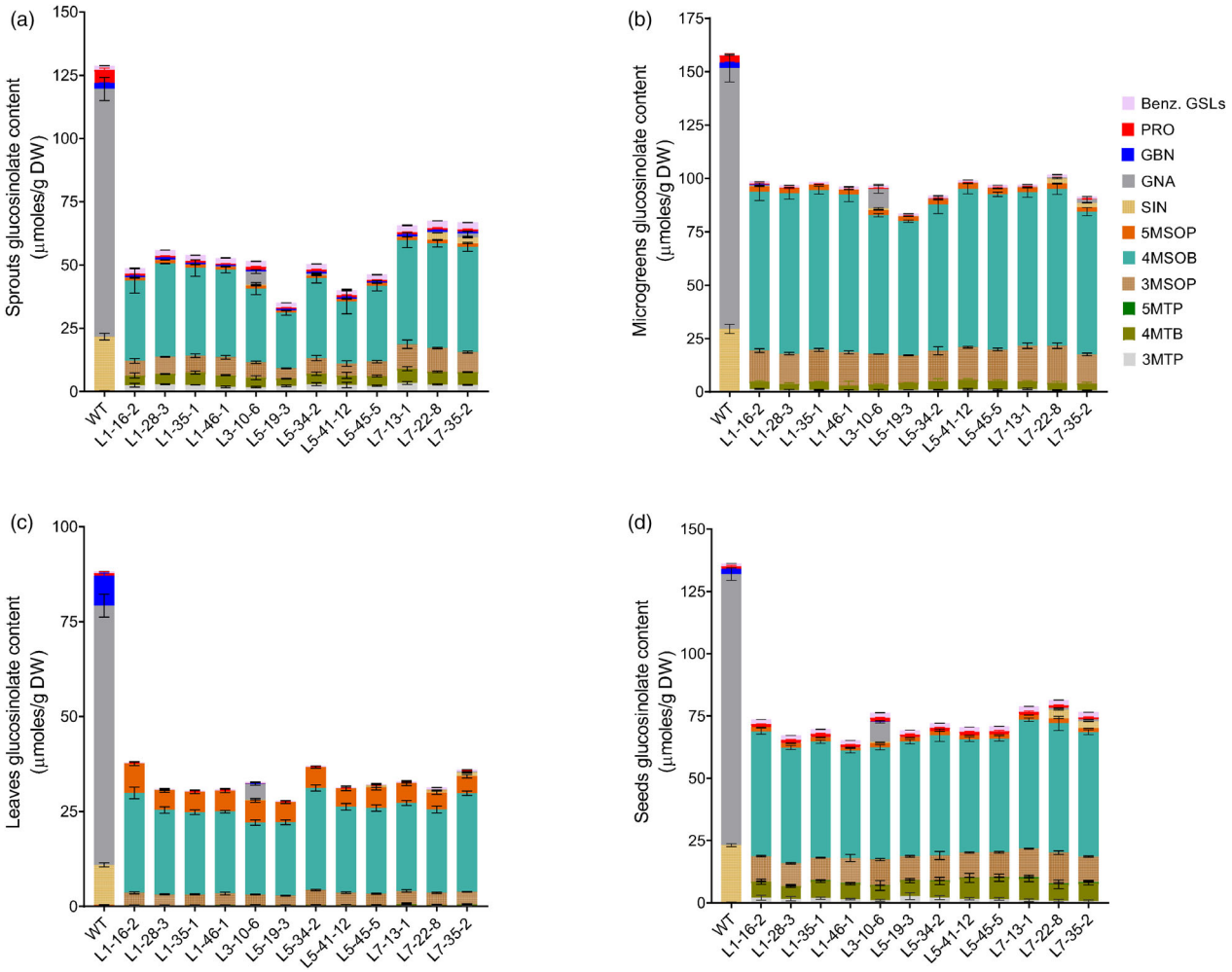

Figure 5 Glucosinolate profiles in different tissue types of 12 transgene-free *BjuAOP2*-edited high glucoraphanin mustard lines in the advanced T_3 generation. Bar plot showing different glucosinolate fractions (µmoles/g DW) in (a) 2 days grown sprouts, (b) 5 days grown microgreens, (c) young leaves of 30–35 days old plants and (d) mature T_4 seeds of the transgene-free *BjuAOP2*-edited T_3 lines and wild-type (WT) Varuna. Each bar represents the mean value ±SE estimated using HPLC ($n > 6$). Data on total and individual glucosinolate content is provided in Tables S9–S12. 3MSOP, 3-methylsulfinylpropyl; 3MTP, 3 methylthiopropyl; 4MSOB, 4-methylsulfinylbutyl (glucoraphanin); 4MTB, 4-methylthiobutyl; 5MSOP, 5-methylsulfinypentyl; 5MTP, 5-methylthiopentyl; GBN, 4-pentenyl (glucobrassicanapin); GNA, 3-butenyl (gluconapin); PRO, 2-hydroxy-3-butenyl (progoitrin); SIN, 2-propenyl (sinigrin); Benz. GSLs include both indolic and benzenic glucosinolates.

µmoles/g DW), microgreens (3.0–5.71 µmoles/g DW), young leaves (0.29–0.61 µmoles/g DW) and seeds (6.88–10.33 µmoles/g DW) of the *BjuAOP2*-edited lines (Figure 5, Tables S9–S12). Notably, a significant reduction of the anti-nutritional alkenyl and hydroxyalkenyl glucosinolates was observed in the *BjuAOP2*-edited lines, leading to a reduction in the total glucosinolates pool. For instance, gluconapin (GNA), the most predominant glucosinolate, was reduced in sprouts (0.0–5.14 µmoles/g DW), microgreens (0–1.73 µmoles/g DW), young leaves (0–4.12 µmoles/g DW) and seeds (0.01–8.09 µmoles/g DW) of the *BjuAOP2*-edited lines compared to 97.92, 128.76, 73.63 and 109.28 µmoles/g DW present in Varuna, respectively (Figure 5, Tables S9–S12). Similarly, sinigrin (SIN), the second major glucosinolate, was also reduced in sprouts (0–2.83 µmoles/g DW), microgreens (0–2.45 µmoles/g DW), young leaves (0–0.92 µmoles/g DW) and seeds (0–3.78 µmoles/g DW) of the *BjuAOP2*-edited lines compared to 21.38, 30.50, 11.23 and 22.47 µmoles/g DW present in Varuna, respectively (Figure 5, Tables S9–S12). In addition, progoitrin (2-hydroxy-3-butenyl, PRO), a key goitrogenic glucosinolate in *Brassica* crops, was also reduced significantly in sprouts (0.65–1.42 µmoles/g DW), microgreens (0.07–0.44 µmoles/g DW), young leaves (0.01–0.10 µmoles/g DW) and seeds (0.89–1.44 µmoles/g DW) of the *BjuAOP2*-edited lines compared to 5.36, 3.32, 0.64 and 1.17 µmoles/g DW present in the wild-type Varuna, respectively (Figure 5, Tables S9–S12). The non-aliphatic class of glucosinolates, namely, indolic and benzenic, remain unaltered in the *BjuAOP2*-edited lines. Overall, editing of *BjuAOP2* homologues provided a higher accumulation of health-beneficial glucoraphanin with a concomitant reduction of the anti-nutritional and goitrogenic alkenyl glucosinolates in mustard.

BjuAOP2-edited high glucoraphanin lines have no compromise on seed quality and yield parameters

Since *B. juncea* is an oilseed crop, we also tested if editing of *BjuAOP2* homologues has any effects on seed quality and yield traits. Fatty acid compositions, total seed oil and protein content were analysed in T_4 seeds of *BjuAOP2*-edited high-glucoraphanin

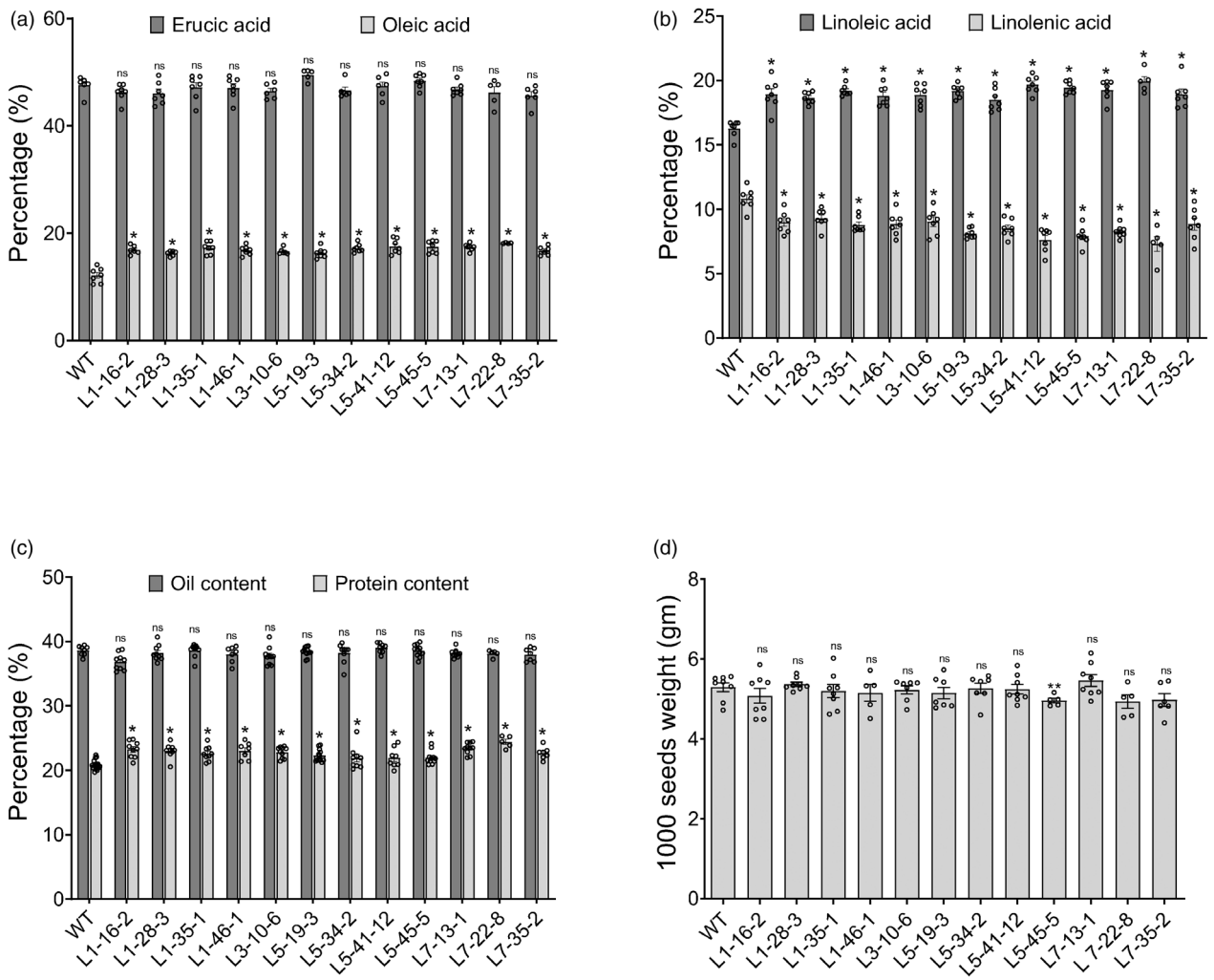

Figure 6 Agronomical performance of 12 transgene-free *BjuAOP2*-edited high glucoraphanin lines in contained growth conditions. Bar plot showing (a) erucic and oleic acid content (in %); (b) linoleic and linolenic acid content (in %); (c) oil and protein content (in %); and (d) 1000 seed weight (in grams) of T_4 generation seeds. The data represents mean ± SE of measurements from 5 to 10 progeny of each transgene-free *BjuAOP2*-edited T_3 line. Asterisks on the top of the bars indicate a significant difference while 'ns' indicates no significant difference between *BjuAOP2*-edited high glucoraphanin lines and wild-type Varuna calculated using two-tailed Student's *t*-test, $P \leq 0.05$. The raw data of various seed quality and yield parameters of *BjuAOP2*-edited lines are provided in Tables S13 and S14.

lines and wild-type Varuna using near infrared reflectance spectroscopy (Table S13). Oil quality is determined by the fatty acid compositions, so different components of fatty acids were also analysed. Oleic acid (C18:1), an important beneficial fatty acid, was found to be significantly increased ($P \leq 0.05$) in *BjuAOP2*-edited high glucoraphanin lines, ranging from 16.24% to 18.14% compared to 12.16% in Varuna (Figure 6a). In addition, linoleic acid (C18:2), linolenic acid (C18:3) and erucic acid (C22:1) ranged between 18.49% and 19.73% ($P \leq 0.05$), 7.62% to 9.29% ($P \leq 0.05$) and 45.61% to 49.31% (non-significant) in T_4 seeds of the *BjuAOP2*-edited lines compared to 16.27%, 10.81% and 47.57% present in the wild-type Varuna, respectively (Figure 6a,b). Total oil content in T_4 seeds of *BjuAOP2*-edited high-glucoraphanin lines ranged from 37.85% to 39.03%, which was at par with Varuna (38.96%) (Figure 6c). Since mustard oilcake serves as a protein-rich animal feed, the total protein content was also analysed. In *BjuAOP2*-edited lines, a significant increase ($P \leq 0.05$) in protein content was observed (21.91%–24.42%) compared to the wild-type Varuna (21.04%) (Figure 6c).

The transgene-free *BjuAOP2*-edited lines were also checked for various agronomical traits during the growing season under the contained net-house conditions. Plant height at maturity, flowering time and various silique parameters such as seeds per pod, silique length and width of the high-glucoraphanin lines were found to be on par with the wild-type Varuna (Tables S14 and S15). In addition, no difference was observed in the seed weight of the high-glucoraphanin lines and Varuna seeds (Figure 6d). Overall, the targeted editing of five *BjuAOP2* homologues allowed us to establish high-glucoraphanin mustard lines, without having any compromise on various seed quality and seed yield-related traits.

Discussion

Glucoraphanin is the major glucosinolate associated with the health-promoting effects in broccoli. Obtaining the beneficial

effects requires an intake of high amounts of broccoli, which has primed a desire to develop the production of a rich source of dietary glucoraphanin. The cultivation and glucoraphanin synthesis in broccoli are dependent on many genetic and environmental factors (Bjorkman *et al*., 2011; Farnham *et al*., 2005), which limits its use as a superfood to a larger population worldwide. In the past, heterologous plant and microbial systems have been used to engineer glucoraphanin biosynthesis. The reconstitution and transient expression of the 13-gene glucoraphanin pathway, localized in different cellular compartments, provided a proof-of-concept to produce glucoraphanin and its precursor dihomomethionine in *Nicotiana benthamiana* (Barnum *et al*., 2022; Crocoll *et al*., 2016; Mikkelsen *et al*., 2010). However, these studies have only been able to produce small amounts of glucoraphanin in heterologous plants compared to the levels present in glucoraphanin-producing plants, suggesting the involvement of accessory pathway genes in the production of health-beneficial glucoraphanin in planta.

Indian mustard (*B. juncea*) is an emerging crop, cultivated worldwide as an oilseed (Cools and Terry, 2018; Frazie *et al*., 2017). In addition, mustard leaves, sprouts and microgreens serve as a healthy substitute for most vegetables as they are rich in total phenolic content and natural anti-oxidants (Li *et al*., 2022; So *et al*., 2023). Mustard oilseed cake is a rich source of proteins and is very much comparable to soy protein, which makes it an alternate suitable candidate for human consumption (Tan *et al*., 2011). However, the potential of *B. juncea* as a superfood crop has not been fully exploited due to high amounts of goitrogenic and anti-nutritional glucosinolates (Augustine and Bisht, 2015; Cools and Terry, 2018; Gohain *et al*., 2021; Mann *et al*., 2023). The presence of functional glucosinolate machinery in *B. juncea* prompted us to engineer high glucoraphanin by editing the *BjuAOP2* gene family, which catalyzes methylsulfinylalkyl to alkenyl glucosinolates.

The allotetraploid *B. juncea* (AB genome) is a natural interspecific hybrid of two species *B. rapa* (A)- and *B. nigra* (B)- with progenitor genomes being paleohexaploids and evolved through whole genome triplication and gene-fractionation events specific to *Brassica* lineage (Augustine *et al*., 2014; Mun *et al*., 2009). As a result, the extant *B. juncea* genome has retained a total of five *BjuAOP2* homologues—representing three 'A' subgenome and two 'B' subgenome homologues. Interestingly, the insertion observed in the second exon of *BjuAOP2.A02*, which led to premature termination, is a unique event in the Indian mustard cv. Varuna, contrary to its counterpart in the progenitor *B. rapa* and the leafy *B. juncea* cv. *tumida* (Chen *et al*., 2023). It seems plausible that during the evolution and domestication of the *B. juncea* divergent gene pools, the homologues of the key glucosinolate pathway genes might have acquired sequence variations to shape the glucosinolate structural diversity (Kumar *et al*., 2019). Despite sharing a high sequence similarity and close evolutionary relationship, the five *BjuAOP2* homologues display distinct expression patterns, wherein the expression dominance of the 'B' subgenome homologues is quite clear in this study. Such expression dominance (or bias) has been reported earlier for glucosinolate pathway genes in *Brassica* species (Augustine *et al*., 2013; Chen *et al*., 2023; Meenu *et al*., 2015). In-depth transcriptomics and genomics studies in recent years have also confirmed the expression dominance of the constituent subgenome(s) as a characteristic feature of many allopolyploids including polyploidy *Brassica* species (Bird *et al*., 2021; Bottani *et al*., 2018; Steige and Slotte, 2016).

In recent years, CRISPR/Cas9-based multiplex genome editing has been utilized for creating genetic lesions in multiple genes/homologues in polyploid *Brassica* crops (Mann *et al*., 2023; Sun *et al*., 2018; Varghese *et al*., 2022). Using this strategy, we also demonstrated that the use of four gRNAs targeting the first exon of the *BjuAOP2* homologues induces frameshift mutations in biallelic homozygous, biallelic heterozygous or chimeric forms, thus ensuring partial to complete loss-of-function of the mutated BjuAOP2 proteins. As a consequence, differential accumulation of seed glucoraphanin content in the *BjuAOP2*-edited lines (5.68%–62.78% of the total glucosinolate content) could be observed in first-generation events. Furthermore, we found that the CRISPR/Cas9-induced *BjuAOP2* mutations segregated independent of the introduced T-DNA fragment (containing *bar* and *Cas9* expression cassettes)—as per the Mendelian norm, generating the transgene-free *BjuAOP2*-edited lines in the advanced generation.

Interestingly, the glucoraphanin accumulation in the transgene-free *BjuAOP2*-edited lines was found to be dose-dependent. For instance, the null mutations in 4–5 *BjuAOP2* homologues could provide 65.69%–71.75% glucoraphanin pool, whereas mutations in 1–3 *BjuAOP2* homologues could provide glucoraphanin only up to 32.41% of the total seed glucosinolate content. Such a dose-dependent effect of homologues has been reported recently in allotetraploids *B. juncea* and *B. napus*. Mann *et al*. (2023) observed that simultaneous editing of 8–10 glucosinolate transporters (*GTR1* and *GTR2*) homologues is needed to generate Canola quality low seed glucosinolate (<30 μmoles/g DW) mustard, whereas editing of 1–7 *GTR* homologues provided a marginal reduction of seed glucosinolates. Similarly, the loss-of-function of up to four *GTR2* homologues is not sufficient to generate the Canola quality lines in *B. napus* (He *et al*., 2022; Tan *et al*., 2022).

In polyploids, homologues often undergo a division of labour by retaining different sub-functions of their original ancestral function (Adams *et al*., 2003; Cheng *et al*., 2014). A recent biochemical study in *B. juncea* cv. *tumida* observed that the five *BjuAOP2* proteins can catalyse the conversion of glucoiberin (or 3MSOP) to sinigrin with different catalytic activities (Chen *et al*., 2023). While sub-functionalization of *BjuAOP2* homologues was quite evident in our study, a few homologues seem to contribute more toward the catalysis of methylsulfinylalkyl (glucoraphanin) to alkenyl (gluconapin) glucosinolates. For instance, the loss-of-function mutations in *BjuAOP2.A09* and *BjuAOP2.B01* homologues exhibited the most prominent effect on glucoraphanin accumulation, compared to *BjuAOP2.A02* mutations causing no further enhancement. Notably, *BjuAOP2.B01* and *BjuAOP2.B03* are the two highly expressed homologues in all the tested tissue types, while *BjuAOP2.A09*, with a comparably lower transcript abundance, also seems to have redundant function for glucoraphanin biosynthesis in *B. juncea*. Collectively, our study provides experimental evidence that homologous genes in allopolyploids undergo expression and functional subfunctionalization, allowing the allopolyploid to differentially use the homologues for variable functions in plant metabolism or in responses to an array of stressful conditions.

Very recently, editing of *AOP2* genes in *B. oleracea* Chinese kale led to the generation of lines with glucoraphanin content of 0.08–0.29 μmoles/g FW (Zheng *et al*., 2023). Using the CRISPR/Cas9-based multiplex editing of *BjuAOP2* homologues allowed us to generate transgene-free mustard lines accumulating glucoraphanin as high as 75.10 μmoles/g DW (approximately

77% of the total glucosinolate pool) in the microgreens, which to the best of our knowledge is the highest reported to date. As expected, the editing of *BjuAOP2* homologues also led to a higher accumulation of methylthioalkyl glucosinolates with a concomitant reduction of the goitrogenic alkenyl and hydroxyalkenyl glucosinolates. The metabolic engineering for high glucoraphanin content via precise editing of multiple *BjuAOP2* homologues does not compromise the agronomical performance and seed yield-related traits in mustard. While these lines also have a higher protein content and beneficial oleic acid, when tested under the contained net-house conditions, the open-field trials of these lines will offer a real picture.

While breeding efforts in *B. oleracea* cv. *italica* have provided a significantly higher glucoraphanin content in the florets of Benefort'e® broccoli hybrid (21.70–30.10 μmoles/g DW) compared to the wild-type broccoli (0.01–14.97 μmoles/g DW), the breeding efforts are often bottlenecked by the unavailability of genetic variation in the breeding population, linkage drags and the quantitative nature of the target trait (Li *et al*., 2021; Traka *et al*., 2013; Yagishita *et al*., 2019). Enhancement of glucoraphanin content has been previously reported using RNAi-based silencing of *AOP2* genes in *B. juncea* (up to 43.11 μmoles/g seed DW) and *B. napus* (up to 42.6 μmoles/g seed DW) and using antisense suppression in Chinese kale (up to 3.03 μmoles/g seed DW) (Augustine and Bisht, 2015; Liu *et al*., 2012; Qian *et al*., 2015); however, these genetically engineered lines will have to follow stringent and exhaustive regulatory processes before they can be cultivated in the farmer field. The transgene-free crops generated through genome-editing technology fall under the SDN-1 category, having more conducive biosafety and regulatory guidelines worldwide (Wolt and Wolf, 2018), which should ease the adoption of high glucoraphanin mustard in the near future.

Taken together, the genome-edited *B. juncea* lines, with such a high glucoraphanin accumulation, could be a new generation of superfood with potential anti-cancer and chemo-preventive benefits. The sprouts, microgreens, leaves, oil and protein-rich oilcake of the *BjuAOP2*-edited mustard crop will be a very good source of health-promoting compounds for humans and animals. Besides, these lines could also be an easy source of large-scale industrial production of high-value glucoraphanin and sulforaphane food supplements.

Materials and methods

Sequence identification and phylogenetic analysis

The full-length coding DNA sequences of *AOP* homologues [BjuVaA02g002933 (*BjuAOP2.A02*), BjuVaA03g002885 (*BjuAOP2.A03*), BjuVaA09g000223 (*BjuAOP2.A09*), BjuVaB01g005452 (*BjuAOP2.B01*) and BjuVaB03g004507 (*BjuAOP2.B03*)] were retrieved from *B. juncea* cv. Varuna genome assembly (Paritosh *et al*., 2021, 2024) using the *A. thaliana AOP* genes [*AtAOP1* (*At4g03070*), *AtAOP2* (*At4g03060;* NCBI locus id: AF417858) and *AtAOP3* (*At4g03050*)] as query sequences. The *AOP*-like sequences from the progenitors *B. rapa* (Bra034180, Bra000848, Bra018521) and *B. nigra* (BniB042223, BniB044970) were retrieved from the BRAD database (http://brassicadb.cn). The annotation of the identified sequences was also confirmed through a BLAST search in the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The identified cDNA sequences were translated into amino acid sequences using the Lasergene core suite (DNASTAR, Madison, WI, USA). Multiple sequence alignment of the deduced amino acid sequences was generated using ClustalW, and the evolutionary tree was constructed using the maximum likelihood method (Bootstrap consensus tree) in MEGA11 software (Tamura *et al*., 2021). Cis-acting regulatory elements present in 2 kb upstream of the transcription start site of *BjuAOP2* homologues were identified using the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot *et al*., 2002).

Gene expression analysis

Different developmental tissues such as stem and leaf from 1-month-old plants, flowers and siliques (3 and 48 days after pollination) were collected and frozen immediately in liquid nitrogen. Total RNA was isolated using a Spectrum Plant Total RNA Kit (Sigma Aldrich, Burlington, MA, USA). Approximately 2.0 μg of RNA was taken for cDNA synthesis using the High capacity cDNA reverse transcription kit (Applied Biosystems, Waltham, MA, USA). RT-qPCR analysis of the 1:20 diluted cDNA samples was performed using gene-specific primers (Table S16) and iTaq universal SYBR Green supermix (Bio-Rad), according to the manufacturer's instructions, in CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA). The *B. juncea actin* gene was used as an endogenous control (Chandna *et al*., 2012). Three independent biological samples, with two technical replicates each, were analysed to derive the conclusions.

Selection of gRNAs and generation of *BjuAOP2* editing construct

The genomic structure of *BjuAOP2* homologues was performed using the gene structure display server (https://gsds.gao-lab.org). Genomic sequences of *BjuAOP2* homologues were used as queries to identify 20 nucleotide gRNAs using CRISPR-P v2.0 (http://crispr.hzau.edu.cn/CRISPR2/). A total of four sgRNAs targeting all the five *BjuAOP2* homologues were selected based on on-score values, exon position and GC content as per *Streptococcus pyrogenes* Cas9 (SpCas9) requirement (Figure S5). To generate the individual sgRNA cassette, a 20 nt gRNA sequence was introduced in between the AtU6-26 promoter and SpCas9 scaffold by a two-step PCR reaction using customized primers (Table S16). In the first step, two independent PCR reactions were performed to get promoter-gRNA and gRNA-scaffold amplicons. In the second step, an overlapping PCR reaction was performed using both amplicons as templates for the generation of ptAtU6-26:gRNA:scaffold expression cassette (Figure S6). The pPZP200 binary vector, containing *SpCas9* driven by Cassava Vein Mosaic Virus (CsVMV) promoter and the plant selection marker '*bar*' gene driven by CaMV35S promoter, was used for cloning all four sgRNA (ptAtU6-26:gRNA:scaffold) cassettes to generate the final *BjuAOP2*(GEd) construct (Figure 2b).

Genetic transformation of *B. juncea* and development of *BjuAOP2*-edited lines

Genetic transformation of a low glucoraphanin *B. juncea* cv. Varuna was performed as per the *Agrobacterium*-mediated transformation protocol established in the laboratory (Augustine *et al*., 2013). In brief, *B. juncea* cv. Varuna seeds were washed properly with a few drops of Teepol B-300 liquid cleaner followed by 70% ethanol and 0.05% $HgCl_2$ treatments. Five-day-old hypocotyls, grown at 23 ± 1 °C, were cut into ~0.5 cm explants and incubated in MS media containing 1 mg/L each of NAA and BAP (N1B1 media) for 24 h in the shaking condition (110 rpm). *Agrobacterium tumefaciens* strain (GV3101) containing *BjuAOP2*

(GEd) construct was grown in YEB liquid medium at 28 °C with appropriate antibiotics and a freshly grown culture of around 0.5–0.6 OD_{600} was re-suspended in N1B1 media to adjust the cell density at ~0.3 OD_{600}. Pre-cultured explants were incubated with *Agrobacterium* cell suspension for 30 min in the dark and co-cultivated further in fresh N1B1 media at 110 rpm for 16–24 h. Co-cultivated explants were washed with N1B1 media containing Augmentin (200 mg/L) and explants were finally plated on N1B1 agar plate containing 10 mg/L 'Basta' (active ingredient phosphinothricin, Bayer AG, Monheim am Rhein, Germany), 20 μM $AgNO_3$ and Augmentin. The regenerated shoots were transferred into the rooting medium containing 2.0 mg/L IBA, 200 mg/L Augmentin and 10 mg/L 'Basta'.

The well-rooted T_0 independent transformants were directly planted into the soil in a containment net-house at the NIPGR experimental field during the growing season (November–March) as per the biosafety guidelines prescribed by the Department of Biotechnology, Government of India. True transgenic events growing in the field were confirmed by 'Basta' leaf painting assay (200 mg/L phosphinothricin). The unopened flower buds were bagged, and self-pollinated seeds were harvested.

Genomic DNA isolation and screening of mutations

Genomic DNA was isolated from the young leaves using the modified CTAB method (Porebski *et al*., 1997). Approximately 500 bp of the genomic sequences flanking the target sites were amplified using 100 ng of genomic DNA as the template, *BjuAOP2* homologues specific primers (Table S16) and ExTaq DNA polymerase (TaKaRa Bio Inc., Kusatsu, Shiga, Japan). The PCR amplicons were gel-eluted and subjected to Sanger sequencing. Chromatograms obtained from different transgenic events were compared with wild-type Varuna using the ICE analysis tool (Synthego Corporation, Redwood City, CA, USA) to identify the mutations in the target sites. Mutation frequency was calculated based on the percent indels of target homologues in transgenic events. Based on the allelic status of edited homologues, transgenic lines were categorized into biallelic homozygous (BHo), biallelic heterozygous (BHt), monoallelic (Mo) and chimeric (Chi).

Identification of transgene-free *BjuAOP2*-edited lines

The self-pollinated T_2 plants from each transgenic event were grown in contained net-house conditions, and true leaves from 20- to 30-day plants were painted with 'Basta' (200 mg/L active ingredient phosphinothricin) for sensitivity assay as described earlier (Mann *et al*., 2023). The segregation of progenies for Basta-resistant (*Bar/Cas9* positive) and Basta-sensitive (*bar/Cas9* negative or transgene-free) lines was recorded after 1 week. The Basta-sensitive lines were also confirmed by *bar*-specific PCR amplification (Figure S7).

Glucosinolate estimation using HPLC

Total glucosinolates were extracted using the protocol established in our laboratory (Augustine and Bisht, 2015). In brief, glucosinolates were extracted from freeze-dried samples (10–20 mg leaf or seeds) in 70% methanol containing 50 μM sinalbin as an internal control. Desulfation of glucosinolate was performed overnight using sulfatase (25 mg/mL, Sigma-Aldrich) on a DEAE Sephadex-A25 column. The desulfoglucosinolates were eluted with 1 mL of HPLC grade water and 5 μL eluent was analyzed with a Shim-pack GIST C18 reverse-phase column (4.6 × 250 mm; 5 μm i.d.) in a Shimadzu CLASS-VP V6.14 HPLC machine. Analysis was done with a gradient of water (solvent A) and 1%–19% (v/v) acetonitrile (solvent B) for 25 min with a flow rate of 1 mL/min and detection of the peak was done at 229 nm. The concentration of glucosinolate components was determined with a known internal standard (sinalbin). The data were plotted using GraphPad Prism 9 (https://www.graphpad.com/).

Performance of high glucoraphanin lines for agronomical, seed yield and seed quality parameters

High glucoraphanin lines were grown in the field under the containment net-house conditions and the plant growth and agronomical parameters were determined at maturity as described previously in our laboratory (Augustine *et al*., 2013; Mann *et al*., 2023). The parameters include plant height, seeds per pod, silique width and length, and thousand seed weight. In addition, various seed quality traits such as total protein, oil content and fatty acid composition (oleic acid, linoleic acid, linolenic acid and erucic acid) of mature dried seeds were determined using near infrared reflectance spectroscopy (NIRS DS-2500, FOSS, Hillerød, Denmark) as described earlier (Augustine and Bisht, 2015; Gohain *et al*., 2021).

Acknowledgements

The work was supported by the BT/PR23893/GET/119/81/2017 and BT/PR53959/PBN/18/17/2024 grants of DBT (India) to N.C.B. and research fellowships to P.K. from the DBT grant and NIPGR short-term fellowship. We thank Avni Mann, Seema Parveen and Mohan Varghese for their help in data presentation, sequence analysis, handling HPLC and their inputs on the manuscript. Central facilities of BRIC-NIPGR and technical help from Amal Roul, Vinod and Raju Das are greatly acknowledged.

Conflict of interest

The authors declare that no conflict of interest exists.

Author contributions

N.C.B. planned and supervised the research work. P.K. performed the experiments, collated the data and drafted the manuscript. All authors read and approved the final manuscript. N.C.B. agrees to serve as the author responsible for contact and ensuring communication.

Data availability statement

The data that supports the findings of this study are available in the supplementary material of this article.

References

Adams, K.L., Cronn, R., Percifield, R. and Wendel, J.F. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl Acad. Sci. USA*, **100**, 4649–4654.

Augustine, R. and Bisht, N.C. (2015) Biofortification of oilseed *Brassica juncea* with the anticancer compound glucoraphanin by suppressing *GSL-ALK* gene family. *Sci. Rep.* **5**, 18005.

Augustine, R. and Bisht, N.C. (2017) Regulation of glucosinolate metabolism: from model plant *Arabidopsis thaliana* to *Brassica* crops. In *Reference Series in Phytochemistry: Glucosinolates*(Ramawat, K. and Mérillon, J.-M., eds), pp. 163–199. Germany: Springer Press.

Augustine, R., Mukhopadhyay, A. and Bisht, N.C. (2013) Targeted silencing of BjMYB28 transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. *Plant Biotechnol. J.* **11**, 855–866.

Augustine, R., Arya, G.C., Nambiar, D.M., Kumar, R. and Bisht, N.C. (2014) Translational genomics in *Brassica* crops: challenges, progress, and future prospects. *Plant Biotechnol. Rep.* **8**, 65–81.

Baralić, K., Živanović, J., Marić, Đ., Bozic, D., Grahovac, L., Miljaković, E.A., Ćurčić, M. *et al*. (2024) Sulforaphane-A compound with potential health benefits for disease prevention and treatment: insights from pharmacological and toxicological experimental studies. *Antioxidants (Basel)*, **13**, 147.

Barnum, C.R., Endelman, B.J., Ornelas, I.J., Pignolet, R.M. and Shih, P.M. (2022) Optimization of heterologous glucoraphanin production in planta. *ACS Synth. Biol.* **11**, 1865–1873.

Bird, K.A., Niederhuth, C.E., Ou, S., Gehan, M., Pires, J.C., Xiong, Z., VanBuren, R. *et al*. (2021) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytol.* **230**, 354–371.

Bjorkman, M., Klingen, I., Birch, A.N.E., Bones, A.M., Bruce, T.J.A., Johansen, T.J., Meadow, R. *et al*. (2011) Phytochemicals of Brassicaceae in plant protection and human health – influences of climate, environment and agronomic practice. *Phytochemistry*, **72**, 538–556.

Blažević, I., Montaut, S., Burčul, F., Olsen, C.E., Burow, M., Rollin, P. and Agerbirk, N. (2020) Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry*, **169**, 112100.

Bottani, S., Zabet, N.R., Wendel, J.F. and Veitia, R.A. (2018) Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* **23**, 393–402.

Cai, C., de Vos, R.C.H., Qian, H., Bucher, J. and Bonnema, G. (2024) Metabolomic and transcriptomic profiles in diverse *Brassica oleracea* crops provide insights into the genetic regulation of glucosinolate profiles. *J. Agric. Food Chem.* **72**, 16032–16044.

Chandna, R., Augustine, R. and Bisht, N.C. (2012) Evaluation of candidate reference genes for gene expression normalization in *Brassica juncea* using real time quantitative RT-PCR. *PLoS One*, **7**, e36918.

Chen, B., Liu, Y., Xiang, C.F., Zhang, D.D., Liu, Z.Y., Liu, Y.H. and Chen, J.J. (2023) Identification and *in-vitro* enzymatic activity analysis of the *AOP2* gene family associated with glucosinolate biosynthesis in tumorous stem mustard (*Brassica juncea* var. *tumida*). *Front. Plant Sci.* **14**, 1111418.

Cheng, F., Wu, J. and Wang, X. (2014) Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.* **1**, 14024.

Clarke, J.D., Dashwood, R.H. and Ho, E. (2008) Multi-targeted prevention of cancer by sulforaphane. *Cancer Lett.* **269**, 291–304.

Cools, K. and Terry, L.A. (2018) The effect of processing on the glucosinolate profile in mustard seed. *Food Chem.* **252**, 343–348.

Crocoll, C., Mirza, N., Reichelt, M., Gershenzon, J. and Halkier, B.A. (2016) Optimization of engineered production of the glucoraphanin precursor dihomomethionine in *Nicotiana benthamiana*. *Front. Bioeng. Biotechnol.* **4**, 14.

Fahey, J.W., Zhang, Y. and Talalay, P. (1997) Broccoli sprouts: an exceptionally rich source of inducers of enzymes that protect against chemicalcarcinogens. *Proc. Natl Acad. Sci. USA*, **94**, 10367–10372.

Fahey, J.W., Wehage, S.L., Holtzclaw, W.D., Kensler, T.W., Egner, P.A., Shapiro, T.A. and Talalay, P. (2012) Protection of humans by plant glucosinolates: efficiency of conversion of glucosinolates to isothiocyanates by the gastrointestinal microflora. *Cancer Prev. Res.* **5**, 603–611.

Farnham, M.W., Stephenson, K.K. and Fahey, J.W. (2005) Glucoraphanin level in broccoli seed is largely determined by genotype. *Hortscience*, **40**, 50–53.

Frazie, M.D., Kim, M.J. and Ku, K.-M. (2017) Health-promoting phytochemicals from 11 mustard cultivars at baby leaf and mature stages. *Molecules*, **22**, 1749.

Gao, M., Li, G., Yang, B., McCombie, W.R. and Quiros, C.F. (2004) Comparative analysis of a *Brassica* BAC clone containing several major aliphatic glucosinolate genes with its corresponding *Arabidopsis* sequence. *Genome*, **47**, 666–679.

Giamoustaris, A. and Mithen, R. (1996) Genetics of aliphatic glucosinolates IV side-chain modification in *Brassica oleracea*. *Theor. Appl. Genet.* **93**, 1006–1010.

Gohain, B., Kumar, P., Malhotra, B., Augustine, R., Pradhan, A.K. and Bisht, N.C. (2021) A comprehensive Vis-NIRS equation for rapid quantification of seed glucosinolate content and composition across diverse *Brassica* oilseed chemotypes. *Food Chem.* **354**, 129527.

Halkier, B.A. and Gershenzon, J. (2006) Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* **57**, 303–333.

Hanlon, N., Coldham, N., Gielbert, A., Kuhnert, N., Sauer, M.J., King, L.J. and Loannides, C. (2008) Absolute bioavailability and dose-dependent pharmacokinetic behaviour of dietary doses of the chemopreventive isothiocyanate sulforaphane in rat. *Br. J. Nutr.* **99**, 559–564.

He, Y., Yang, Z., Tang, M., Yang, Q.Y., Zhang, Y. and Liu, S. (2022) Enhancing canola breeding by editing a glucosinolate transporter gene lacking natural variation. *Plant Physiol.* **188**, 1848–1851.

Higdon, J.V., Delage, B., Williams, D.E. and Dashwood, R.H. (2007) Cruciferous vegetables and human cancer risk: epidemiologic evidence and mechanistic basis. *Pharmacol. Res.* **55**, 224–236.

Hopkins, R.J., van Dam, N.M. and van Loon, J.J.A. (2009) Role of glucosinolates in insect plant relationships and multitrophic interactions. *Annu. Rev. Entomol.* **54**, 57–83.

Houghton, C.A. (2019) Sulforaphane: its "Coming of Age" as a clinically relevant nutraceutical in the prevention and treatment of chronic disease. *Oxidative Med. Cell. Longev.*, 2716870.

Kaiser, A.E., Baniasadi, M., Giansiracusa, D., Giansiracusa, M., Garcia, M. and Fryda, Z. (2021) Sulforaphane: a broccoli bioactive phytocompound with cancer preventive potential. *Cancers (Basel)*, **13**, 4796.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J. and Mitchell-Olds, T. (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell*, **13**, 681–693.

Kumar, R., Lee, S.G., Augustine, R., Reichelt, M., Vassão, D.G., Palavalli, M.H., Allen, A. *et al*. (2019) Molecular basis of the evolution of methylthioalkylmalate synthase and the diversity of methionine-derived glucosinolates. *Plant Cell*, **31**, 1633–1647.

Kushad, M.M., Brown, A.F., Kurilich, A.C., Juvik, J.A., Klein, B.P., Wallig, M.A. and Jeffery, E.H. (1999) Variation of glucosinolates in vegetable crops of *Brassica oleracea*. *J. Agric. Food Chem.* **47**, 1541–1548.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. *et al*. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327.

Li, Z., Zheng, S., Liu, Y., Fang, Z., Yang, L., Zhang, M., Zhang, Y. *et al*. (2021) Characterization of glucosinolates in 80 broccoli genotypes and different organs using UHPLC-Triple-TOF-MS method. *Food Chem.* **334**, 127519.

Li, Z., Di, H., Cheng, W., Ren, G., Zhang, Y., Ma, J., Ma, W. *et al*. (2022) Effect of the number of dark days and planting density on the health-promoting phytochemicals and antioxidant capacity of mustard (*Brassica juncea*) sprouts. *Plants (Basel)*, **11**, 2515.

Liu, Z., Hirani, A.H., McVetty, P.B., Daayf, F., Quiros, C.F. and Li, G. (2012) Reducing progoitrin and enriching glucoraphanin in *Brassica napus* seeds through silencing of the *GSL-ALK* gene family. *Plant Mol. Biol.* **79**, 179–189.

Liu, P., Zhang, B., Li, Y. and Yuan, Q. (2024) Potential mechanisms of cancer prevention and treatment by sulforaphane, a natural small molecule compound of plant-derived. *Mol. Med.* **30**, 94.

Mann, A., Kumari, J., Kumar, R., Kumar, P., Pradhan, A.K., Pental, D. and Bisht, N.C. (2023) Targeted editing of multiple homologues of *GTR1* and *GTR2* genes provides the ideal low-seed, high-leaf glucosinolate oilseed mustard with uncompromised defence and yield. *Plant Biotechnol. J.* **21**, 2182–2195.

Meenu, Augustine, R., Majee, M., Pradhan, A.K. and Bisht, N.C. (2015) Erratum to: Genomic origin, expression differentiation and regulation of multiple genes encoding CYP83A1, a key enzyme for core glucosinolate biosynthesis, from the allotetraploid *Brassica juncea*. *Planta*, **241**, 667.

Mikkelsen, M.D., Olsen, C.E. and Halkier, B.A. (2010) Production of the cancer-preventive glucoraphanin in tobacco. *Mol. Plant*, **3**, 751–759.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H. *et al*. (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* **10**, R111.

Østergaard, L. and King, G.J. (2008) Standardized gene nomenclature for the *Brassica* genus. *Plant Methods*, **4**, 10.

Paritosh, K., Yadava, S.K., Singh, P., Bhayana, L., Mukhopadhyay, A., Gupta, V., Bisht, N.C. *et al*. (2021) A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *Plant Biotechnol. J.* **19**, 602–614.

Paritosh, K., Rajarammohan, S., Yadava, S.K., Sharma, S., Verma, R., Mathur, S., Mukhopadhyay, A. *et al*. (2024) A chromosome-scale assembly of *Brassica carinata* (BBCC) accession HC20 containing resistance to multiple pathogens and an early generation assessment of introgressions into *B. juncea* (AABB). *Plant J.* **119**, 762–782.

Porebski, S., Bailey, L.G. and Baum, B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Report.* **15**, 8–15.

Qian, H.M., Sun, B., Miao, H.Y., Cai, C.X., Xu, C.X. and Wang, Q.M. (2015) Variation of glucosinolates and quinone reductase activity among different varieties of Chinese kale and improvement of glucoraphanin by metabolic engineering. *Food Chem.* **168**, 321–326.

Russo, M., Spagnuolo, C., Russo, G.L., Skalicka-Woźniak, K., Daglia, M., Sobarzo-Sánchez, E., Nabavi, S.F. *et al*. (2018) Nrf2 targeting by sulforaphane: a potential therapy for cancer treatment. *Crit. Rev. Food Sci. Nutr.* **58**, 1391–1405.

Sivapalan, T., Melchini, A., Saha, S., Needs, P.W., Traka, M.H., Tapp, H., Dainty, J.R. *et al*. (2018) Bioavailability of glucoraphanin and sulforaphane from high-glucoraphanin broccoli. *Mol. Nutr. Food Res.* **62**, e1700911.

So, V., Poul, P., Oeung, S., Srey, P., Mao, K., Ung, H., Eng, P. *et al*. (2023) Bioactive compounds, antioxidant activities, and HPLC analysis of nine edible sprouts in Cambodia. *Molecules*, **28**, 2874.

Steige, K.A. and Slotte, T. (2016) Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Curr. Opin. Plant Biol.* **30**, 88–93.

Sun, Q., Lin, L., Liu, D., Wu, D., Fang, Y., Wu, J. and Wang, Y. (2018) CRISPR/Cas9-mediated multiplex genome editing of the *BnWRKY11* and *BnWRKY70* genes in *Brassica napus* L. *Int. J. Mol. Sci.* **19**, 2716.

Tamura, K., Stecher, G. and Kumar, S. (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027.

Tan, S.H., Mailer, R.J., Blanchard, C.L. and Agboola, S.O. (2011) Canola proteins for human consumption: extraction, profile, and functional properties. *J. Food Sci.* **76**, R16–R28.

Tan, Z., Xie, Z., Dai, L., Zhang, Y., Zhao, H., Tang, S., Wan, L. *et al*. (2022) Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in *Brassica napus*. *Plant Biotechnol. J.* **20**, 211–225.

Traka, M.H., Saha, S., Huseby, S., Kopriva, S., Walley, P.G., Barker, G.C., Moore, J.D. *et al*. (2013) Genetic regulation of glucoraphanin accumulation in Beneforte® broccoli. *New Phytol.* **198**, 1085–1095.

Varghese, M., Malhotra, B. and Bisht, N.C. (2022) Genome editing in polyploid *Brassica* crops. In *The Brassica Juncea Genome. Compendium of Plant Genomes*(Kole, C. and Mohapatra, T., eds), pp. 471–491. Cham: Springer.

Verkerk, R., Schreiner, M., Krumbein, A., Ciska, E., Holst, B., Rowland, I., De Schrijver, R. *et al*. (2009) Glucosinolates in *Brassica* vegetables: the influence of the food supply chain on intake, bioavailability and human health. *Mol. Nutr. Food Res.* **53**, S219–S265.

Wang, R., Ripley, V.L. and Rakow, G. (2007) Pod shatter resistance evaluation in cultivars and breeding lines of *Brassica napus*, *B. juncea* and *Sinapis alba*. *Plant Breed.* **126**, 588–595.

Wolt, J.D. and Wolf, C. (2018) Policy and governance perspectives for regulation of genome-edited crops in the United States. *Front. Plant Sci.* **9**, 1606.

Yagishita, Y., Fahey, J.W., Dinkova-Kostova, A.T. and Kensler, T.W. (2019) Broccoli or sulforaphane: is It the source or dose that matters? *Molecules*, **24**, 3593.

Zandberg, J.D., Fernandez, C.T., Danilevicz, M.F., Thomas, W.J.W., Edwards, D. and Batley, J. (2022) The global assessment of oilseed *Brassica* Crop species yield, yield stability and the underlying genetics. *Plan. Theory*, **11**, 2740.

Zhang, Y., Talalay, P., Cho, C.G. and Posner, G.H. (1992) A major inducer of anticarcinogenic protective enzymes from broccoli: Isolation and elucidation of structure. *Proc. Natl Acad. Sci. USA*, **89**, 2399–2403.

Zhang, Y., Zhang, W., Zhao, Y., Peng, R., Zhang, Z., Xu, Z., Simal-Gandara, J. *et al*. (2024) Bioactive sulforaphane from cruciferous vegetables: advances in biosynthesis, metabolism, bioavailability, delivery, health benefits, and applications. *Crit. Rev. Food Sci. Nutr.* **65**, 1–21.

Zheng, H., Huang, W., Li, X., Huang, H., Yuan, Q., Liu, R., Di, H. *et al*. (2023) CRISPR/Cas9-mediated *BoaAOP2s* editing alters aliphatic glucosinolate side chain metabolic flux and increases the glucoraphanin content in Chinese kale. *Food Res. Int.* **170**, 112995.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 General scheme for aliphatic glucosinolate biosynthesis.
Figure S2 Multiple sequence alignment of *BjuAOP2* homologues with *Arabidopsis AOP2* genomic and CDS.
Figure S3 Sequence identity of coding sequences of *BjuAOP2* homologues with *Arabidopsis AOP* genes.
Figure S4 Cis-acting regulatory elements in the promoter region of *BjuAOP2* homologues.
Figure S5 Location of the four guide RNAs targeting the *BjuAOP2* homologues.
Figure S6 Generation of sgRNA through two-step PCR strategy.
Figure S7 Identification of transgene-free *BjuAOP2*-edited lines.
Table S1 Comparison of gene structure of *B. juncea BjuAOP2* homologues.
Table S2 Summary of the different gRNAs used for the development of *BjuAOP2*(GEd) construct.
Table S3 Estimation of fractions and total seed glucosinolate (μmoles/g DW) in the T_1 seeds of *BjuAOP2*-edited events and the wild-type (WT) Varuna.
Table S4 Genotype of CRISPR/Cas9-induced mutations in the *BjuAOP2* homologues in the *BjuAOP2*-edited T_0 events.
Table S5 Mutation frequency recorded in *BjuAOP2* homologues in four independent *BjuAOP2*-edited T_0 events.
Table S6 Summary of transgene-free T_2 plants obtained for each transformation event derived from *BjuAOP2*(GEd) construct.
Table S7 Estimation of total and fraction glucosinolates (μ moles/g DW) in the T_3 seeds of *BjuAOP2*-edited lines.
Table S8 Genotype of CRISPR/Cas9-induced mutations of *BjuAOP2* homologues in the selected *BjuAOP2*-edited transgene-free T_2 lines.
Table S9 Estimation of different fractions and total glucosinolate content (μmoles/g DW) in the 2 days old sprouts of *BjuAOP2*-edited lines and the wild-type (WT) Varuna.
Table S10 Estimation of different fractions and total glucosinolate content (μmoles/g DW) in the 5 days old microgreen of *BjuAOP2*-edited lines and the wild-type (WT) Varuna.
Table S11 Estimation of total and fractional glucosinolate content (μmoles/g DW) in the young leaves of transgene-free *BjuAOP2*-edited T_3 lines.
Table S12 Estimation of total and fractional glucosinolate content (μmoles/g DW) in the T_4 seeds of *BjuAOP2*-edited lines.
Table S13 Seed quality parameters in T_4 seeds of 12 high glucoraphanin *BjuAOP2* lines and WT Varuna seeds estimated using near-infrared spectroscopy.
Table S14 Agronomical parameter of 12 high glucoraphanin *BjuAOP2* lines.
Table S15 Comparison of days to flowering between wild-type Varuna and high glucoraphanin *BjuAOP2*-edited lines.
Table S16 List of primers used in this study.