### RESEARCH



# The development of ideal insertion and deletion (InDel) markers and initial indel map variation in cucumber using resequenced data

Jinchang Yang<sup>1,2†</sup>, Pengxu Meng<sup>2†</sup>, Hang Mi<sup>2†</sup>, Xueying Wang<sup>2</sup>, Jin Yang<sup>1,3</sup>, Shaohong Fu<sup>1,3</sup>, Wanzhuo Gong<sup>1,3</sup>, Ruifan Bao<sup>2</sup>, Wujiao Deng<sup>2</sup>, Haimei Wu<sup>2</sup>, Zhiqin Liu<sup>2</sup>, Sijia Yang<sup>2</sup>, Jixing Ni<sup>2</sup>, Jian Liu<sup>2\*</sup> and Haoran Shi<sup>1,3\*</sup>

#### Abstract

**Background** InDels are the most common type of length polymorphism and play a critical role in the genetic traits of many important phenotypes in both plants and animals, making them an ideal source for length polymorphism molecular markers. However, in the process of cucumber breeding, researchers still face deficiencies in the identification of InDel loci and the development of genomic-wide molecular markers.

**Results** In this study, we conducted InDel identification on 115 cucumber re-sequencing datasets, identifying a total of 7,842,946 InDels, with lengths ranging from 1 to 59 bp and an average density of one InDel every 2.8 kb on the chromosomes. The InDel variations were classified into four main categories, and 81 InDel hotspots were identified, serving as the foundation for constructing a cucumber InDel variation map. Additionally, we utilized an electronic PCR strategy to develop genome-wide InDel markers for cucumber, resulting in the selection of 22,442 InDel primers exhibiting high polymorphism (PIC  $\geq$  0.5) and major allele differences of  $\geq$  3 bp. We experimentally validated 50 randomly selected InDel primers, and the results showed that all markers exhibited high polymorphism.

**Conclusions** The construction of the cucumber genome InDel variation map aids in understanding the genetic basis of key traits in cucumber derived from InDel variations. The ideal InDel markers developed in this study may enhance the efficiency of cucumber breeding for resistance to both biotic and abiotic stresses, as well as scientific research.

Keywords InDels, Cucumber, Genome variation, Re-sequenced, Molecular marker

 $^\dagger \mathrm{Jinchang}\,\mathrm{Yang},$  Pengxu Meng and Hang Mi contributed equally to this work.

\*Correspondence: Jian Liu liujian@sicau.edu.cn Haoran Shi sicau3339@outlook.com <sup>1</sup>Chengdu Academy of Agricultural and Forestry Sciences, No 200 Nongke Road, Wenjiang District, Chengdu, Sichuan 611130, China <sup>2</sup>Maize Research Institute, Sichuan Agricultural University, No 211 Huiming Road, Wenjiang District, Chengdu, Sichuan 611130, China <sup>3</sup>Chengdu Research Branch, National Rapeseed Genetic Improvement Center, No 200 Nongke Road, Wenjiang District, Chengdu, Sichuan 611130, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creative.commons.org/licenses/by-nc-nd/4.0/.

#### Introduction

Cucumber (*Cucumis sativus* L., 2n = 2x = 14) is an annual herbaceous climbing plant and is considered one of the most economically significant vegetable crops worldwide [1]. Currently, China is the top cucumber-producing country, ranking first globally in both the scale of cultivation and yield. The cucumber is also an important model plant for research in vascular plant biology and sex determination and is recognized as a plant diversity center of significant biological value [2]. However, with the continuous improvement of yield and quality in cucumber breeding, the vegetable is beginning to encounter bottlenecks. In addition, in recent years, frequent extreme weather events, droughts, pests and diseases have caused losses to cucumber production worldwide, indicating a need to improve the stress resistance in many traditional varieties. Therefore, molecular marker technology is used to provide scientific means to address these issues and several markers have been applied in stress-related research. For instance, molecular markers such as amplified fragment length polymorphism [3-5], random amplified polymorphic DNA [6, 7], SSR [8, 9], single nucleotide polymorphism (SNP) [10], and InDels are widely used in genetic diversity analysis, gene mapping, and marker-assisted breeding (MAB) of cucumbers. The high polymorphic indel markers were also used to analyze the genetic diversity of cucumber inbred lines [11, 12], leading to fine mapping and gene cloning of important trait genes such as bitterness gene Bt [13] and powdery mildew resistance gene [14]. These laid a foundation for research on cucumber germplasm diversity and for further exploration of important quality genes. However, as research on fine mapping and molecular MAB intensifies, the availability of InDel markers is becoming a limiting factor.

InDels are collections of insertions and deletions of bases ranging from 1 to 50 bp and are the most widespread type of length polymorphism variations in genomes. In plants, InDels provide abundant genetic variation and influence gene expression as well as trait changes, and account for approximately 10-15% of genomic variation [15, 16]. For example, InDel variations can influence crop yield by regulating the number of branches or grain size [17–19]; they can also regulate the synthesis of anthocyanins, thereby changing the color of the fruit [20, 21]; additionally, they can enhance the cold and disease resistance of plants [22, 23]. Therefore, identifying InDel variations in plant genomes is crucial for understanding their impact on trait development. The rapid development of sequencing technologies and the reduced sequencing costs have promoted the production of large-scale resequencing data, resulting in the publication of several human, rice and maize re-sequenced genome data through which InDels can be sourced.

About 20 vegetables and fruits have also been studied using intraspecific resequencing data. These data have led to the identification and characterization of 1.6 million InDels in the human genome [22], and millions and hundreds of thousands of InDels in maize [15] and rice [16], respectively. The rapid development of sequencing technologies and the reduced sequencing costs have promoted the production of large-scale resequencing data, resulting in the publication of several human, rice and maize re-sequenced genome data through which InDels can be sourced. About 20 vegetables and fruits have also been studied using intraspecific resequencing data. These data have led to the identification and characterization of 1.6 million InDels in the human genome [24], and millions and hundreds of thousands of InDels in maize [15] and rice [16], respectively.

The InDel variations provide the highest quality data source for developing high-density length polymorphism molecular markers, named InDel markers. Traditionally, simple sequence repeat (SSR) markers have been the most widely used type of length polymorphism marker in most laboratories due to their low cost, minimal equipment requirements, low technical demands, and high accuracy [25, 26]; however, in maize and rice, the number of SSR markers is significantly smaller than that of InDel markers [15, 16, 27, 28]. Therefore, the high-density character of the InDel markers makes them the most ideal length polymorphism markers. Despite this, there is an increasing demand for high-density InDel markers across different species. However, developing InDel markers from the InDel variations requires several conditions to be met, including distribution at a single locus within the genome, high polymorphism and conservation of flanking sequences, and the length difference suitable for electrophoresis resolution and the amplification length ratio. The large-scale development of InDel molecular markers is also constrained by the need for specialized software skills and computational resources.

Therefore, in this study, we called SNP and InDel variations using GATK in the resequencing data of 115 cucumber samples and used the data to construct a whole-genome InDel variation map for cucumber and analyze their composition and distribution patterns. We also used the previously developed electronic PCR (e-PCR)-based InDel marker development strategy [15, 16] to develop and validate a set of high-density InDel markers. Overall, the established InDel variation map facilitates the identification of polymorphisms that directly affect important cucumber agronomic characters and biological and abiotic stress. This set of developed ideal InDel markers can also enhance the efficiency of both fundamental and applied cucumber research in fine mapping and molecular-assisted breeding (MAB).

#### **Materials and methods**

#### Sequencing data and plant materials of the cucumber

The cucumber reference genome sequences and gene annotation (Cucumber\_ v4, 321.53 Mb) of the inbred line 9930 derived from a Chinese long cultivar in China were downloaded from the website (http://www.cucumberdb. com/) [29]. The genome was annotated into five regions, including 5'-UTR, coding determining sequences (CDS), 3'-UTR, intron and intergenic regions. The 500 bp flanking sequences of the gene region were analyzed as TSS\_UP\_0.5Kb and TES\_down\_0.5 kb, respectively. A total 115 of cucumber re-sequence genome data were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/sra/?te rm=SRA056480) [30].

A set of 39 cucumber lines was chosen for the polymorphism verification, including 19 materials of 115 the re-sequenced cucumber samples (Hw 2, JL-14 Dhillon, 10382, 9779, 13598 SC 53-B [6], 2163, SC 50, GY14, Marketmore76, 9930, CM8537-1-2-1-1-0-1-1-1, He Cha Huang Gua, Qian Qi Li Huang Gua, Jia Huang Gua, Sekino No. 2, Yuan Bai Huang Gua, Bai Pi Yuan Huang Gua, Ban Na Huang Gua 53 and Yuan Zong Huang Di Huang Gua),18 new cucumber lines (14#-27 S, 14#-7, 14#-21 S, 14#-23, 14#-67 S, 14#-9, 14#-13, 11#-X45S, 11#-63 S, 11#-X34, 11#-X83B, 11#-X81, 11#-X45S, 11#-X55, 11#-X57S, 11#-X67S, 11#-X77S and DA40) and two samples wild cucumber (C. sativus var. hardwickii: 08CS986 and 08CS553) are provided by the Tianjin Kernel Cucumber Research Institute.

#### Read mapping and indel discovery using GATK

After filtering the reads of the 115 accessions with a cutoff quality score of 20 using the NGS QC toolkit (v2.3.3), the reads were mapped to the reference genome using Burrows-Wheeler Alignment (BWA) (v0.7.17) with the default parameters. The reads that were mapped to multiple loci in the genome were removed using Picard MarkDuplicates (http://picard.sourceforge.net/comm and-line-overview.shtml). The remaining reads were realigned around InDels using the RealignerTargetCreator and InDelRealigner of Genome Analysis Toolkit (GATK) tools (v4.0.4.0). The InDel variants were called and filtered using the GATK UnifiedGenotyper and VariantFiltration with the default parameters, respectively.

#### Statistical analysis of variation hotspots

The genome was divided into 100 kb bins, and the variation levels were calculated relative to the InDel and SNP numbers in each bin across the population. Mean and standard deviation analyses were conducted for InDel and SNP variation levels for each chromosome, and a one-sided Z-test was performed for each bin. To control for multiple-sampling errors, the resulting P-values were adjusted using false discovery rate (FDR) analysis [31, 32]. A conservative FDR cutoff of (Q-value) < 0.1 was applied, and the results of highly significant InDel and SNP hotspots are provided in Supplementary Table 2, along with corresponding P-values.

#### Assembled contigs

The reads with a quality score of 20 and above were selected for subsequent contig assembly using ABySS [33]. Afterwards, we chose a k-mer size of 45 for 90 bp read length samples and 25 for 44 bp read length samples for the contig assembly of re-sequenced samples. In our previous study [34], we utilized resequencing data from 115 samples with an average depth of 16–30 for contig assembly, resulting in a total of 221,985 to 608,332 overlapping groups. The total number of bases ranged from 56,713,982 to 386,317,213 bp, with an N50 of 70 to 17,416. The maximum overlapping group per sample ranged from 11,551 to 235,183 bp. Additional specific data can be found in Supplementary Table 3.

## Highly polymorphic indel marker identified by an e-PCR strategy

We employed the InDel calling and polymorphism discovery pipeline previously developed for identifying polymorphic InDel markers in rice and maize [15, 16], which comprises four stages (the detailed procedure can be found in Supplementary File 1). First, using the Python script 'step1.written\_in\_one\_line\_and\_primer\_design.py', primers for electronic PCR were designed based on the cucumber reference genome 9930 (CLv4.0) as a template. The primers covered the entire genome, with a primer length of 20 bp and an amplicon length of 60 bp, with a sliding window of 20 bp, ensuring full coverage of the cucumber genome. Then, using the Python script 'step2. primer\_e-PCR.py, the primer sequences were aligned to the cucumber Clv4.0 genome. The Python script 'step3. select\_single\_matching\_primers.py' was then used to filter for unique primers, which have a unique position and sequence in the genome. Finally, these unique primers were used for electronic PCR with contig sequence data from 115 samples, ultimately identifying the InDels. The polymorphism of the primers with a unique genotype in a sample and 20 or more genotypes in the population was then assessed using polymorphism information content (PIC) using the formula.

$$\operatorname{PIC}_{i} = 1 - \sum_{j=1}^{n} p_{ij}^{2}$$

Where pij was the frequency of the jth pattern for the ith markers [35].

#### PCR validation of highly polymorphic indel marker

A 200 bp region, encompassing a 20 bp InDel region with 90 bp flanking sequences on each side, was used to design the primer using Primer5 software. The design parameters were applied: included a primer length of 18-25 nucleotides with an optimal length of 22 nucleotides, a Tm of  $55^{\circ}$ C to  $64^{\circ}$ C with  $60^{\circ}$ C as optimal, an amplicon product length of 60 to 200 bp, and the preference for a G- or C-rich perfect ending at the 3'-end [15, 16].

A set of 50 highly polymorphic InDel markers with a PIC value greater than or equal to 0.5 and major allele differences of at least 3 bp were randomly selected to validate the polymorphism in the DNA extracted from 39 cucumber lines. Selection thresholds were established at

Table 1	Classification	of indels from	115 re-sequenced
cucumb	er samples		

InDel Class	Number	Rate
Single bases	4,481,885	57.15%
A	1,749,883	22.31%
Т	1,742,058	22.21%
С	496,833	6.33%
G	493,111	6.29%
Repeat Expansions	3,118,580	39.76%
Monomeric (4/4)	703,289	8.97%
(T)n	332,623	4.24%
(A)n	331,437	4.23%
(C)n	19,834	0.25%
(G)n	19,395	0.25%
Dimeric (12/12)	1,139,311	14.53%
(AT)n	244,447	3.12%
(TA)n	224,719	2.87%
(AG)n	107,640	1.37%
(TC)n	102,126	1.30%
(TG)n	94,232	1.20%
(AC)n	89,467	1.14%
Other (NN)n	276,680	3.53%
Trimeric (60/60)	678,902	8.66%
(AAG)n	678,902	8.66%
(AAT)n	40,479	0.52%
(TTC)n	35,863	0.46%
(CTT)n	31,313	0.40%
(ATT)n	31,432	0.40%
(AAC)n	25,933	0.33%
Other (NNN)n	513,882	6.53%
Tetrameric (245/252)	144,843	1.85%
Pentameric (1009/1020)	157,633	2.01%
Hexameric (3613/4092)	36,854	0.47%
Heptameric (6813/16380)	103,379	1.32%
Octameric (9002/65532)	69,310	0.88%
Nomameric (9772/262140)	26,537	0.34%
Decameric (9927/1048572)	58,522	0.75%
Transposons	269	0.00%
other	242,212	3.09%
total	7,842,946	

3 bp and 8 bp: primers with allelic differences of  $\geq$  3 bp are employed to design PCR products approximately 60–100 bp in length, typically detected using polyacryl-amide gel electrophoresis; primers with allelic differences of  $\geq$  8 bp are used to develop PCR products approximately 150–300 bp in length, distinguishable via agarose gel electrophoresis.

The 20  $\mu$ L PCR reaction mixture comprising 50 ng DNA, 2.0  $\mu$ L of 10 x buffer (containing Mg<sup>2+</sup>), 2.5  $\mu$ L dNTP (2.5 mM), 100 nM of each primer, 1.5 U Taq polymerase, and ddH<sub>2</sub>O, was amplified as described previously. The amplicons were electrophoresed on a 6% polyacrylamide gel or 2% agarose gel, visualized through silver staining and used to calculate PIC.

#### GO analysis

The GO enrichment analysis of genes associated with highly polymorphic (PIC  $\geq$  0.5) InDel markers was conducted using the Singular Enrichment Analysis approach through the online AgriGO tool (http://bioinfo.cau.edu.cn/agriGO/) with the cucumber reference genome as the background [36]. Highly significant enriched terms were selected based on the default P-values and FDR.

#### Results

#### InDels variations in cucumber genome

A total of 7,842,946 non-redundant InDels were identified from 115 independent sets of reads. The InDels ranged from 1 to 59 bp with a mean of 2.59 bp, and are distributed throughout the seven chromosomes with an average density of one InDel per 24.39 kb. Four major InDel classes were identified, including single-base pair, monomeric base pair expansion, multi-base pair expansion of 2–15 bp repeat unit, and random DNA sequence InDels. The majority of the InDels (57.15%) were singlebase pairs, followed by monomeric base pairs and multibase repeat expansions of various lengths with 39.76% and random DNA sequence InDels with 3.09%. Among the single-base pair InDels, the AT and TA base pairs were the majority comprising 77.90% of these InDels (Table 1, S1).

#### InDel and SNP hotspots

The number of InDels per 100 kb bin exhibited substantial variation across each chromosome, ranging from 172 to 10,300, with an average of 4,012.38. The number of SNPs per 100 kb ranged from 1,255 to 90,619, with an average of 29,115.09 (Fig. 1; Supplementary Table 2). Upon mapping these data to each bin, 81 InDel variation hotspots were identified across all chromosomes, including 38 InDel-only hotspots, with significant levels of variation. The number of InDels and SNPs within all bins on the chromosomes varied greatly, with a single bin sometimes containing 60 and 72 InDels and SNPs, respectively,



Fig. 1 The distributions for SNPs and InDels on cucumber chromosomes. "A" and "B": the distributions for overall SNPs and InDels. The identified DNA variations were summarized as numbers in 1 Mb bins along each of seven chromosomes

more than the lowest counts. Additionally, nearly half of the bins have InDel and SNP counts exceeding the average level, with the top 10% of bins having average counts that are 4.68 and 6.49 times higher than the bottom 10%, respectively. This indicates that these regions represent hotspots of genetic variation.

#### Mining ideal indel markers

Using the e-PCR-based InDel marker development strategy [6, 7], we designed 295,728,002 pairs of e-PCR primers with the cucumber reference genome DNA sequence as the template. Among the primers, 8,953,755 (3.03%) pairs were unique with an average density of 0.92 InDels/ bp across the seven chromosomes and 18.71% (1,675,641) of them were polymorphic (Table 2, Supplementary Table 4). The PIC of the InDels in the cucumber population ranged from 0.01 to 0.95, averaging 0.10 (Fig. 2). The size difference of more than half of the InDel loci was within 3 bp (Fig. 3). The number of alleles generated for all InDels varied from 2 to 28, averaging 2.37, with the majority (80.09%) being bi-allelic mutations (Fig. 4). The

Table 2 Distribution of e-PCR primer and polymorphic indels in different regions of the cucumber genome

Genome Region	Total	Unique Primer <sup>a</sup>	InDels <sup>b</sup> (PIC>0) High polymorphicInDel(PIC $\ge$ 0.5)			.5)		
			Count	Density <sup>c</sup> (number/kp)	Rate <sup>d</sup> (%)	Count	Density (number/kp)	Rate <sup>e</sup> (%)
TSS_up_0.5 kb	53,199,600	589,796	163,160	3.07	27.66	7,522	12.75	1.28
5'UTR	3,357,940	156,913	31,836	9.48	20.29	1,585	10.10	1.01
3'UTR	5,603,815	278,243	54,050	9.65	19.43	2,336	8.40	0.84
CDS	5,603,815	1,585,140	312,348	8.23	19.70	1,902	1.20	0.12
Intron	43,955,360	2,197,768	268,875	6.12	12.23	13,685	6.23	0.62
TES_down_0.5 kb	28,209,168	605,347	109,337	3.88	18.06	7,438	12.29	1.23
Intergenic	123,470,359	3,953,671	815,854	6.61	20.64	53,228	13.46	1.35
Total <sup>f</sup>	295,728,002	8,953,755	1,675,641	5.67	18.71	86,301	9.64	0.96

a. Primers were located in unique genomic region

b. Polymorphisms primers were primers with length information in twenty or more than twenty genomes and PIC Value >0

c. Density was calculated by number/Kp

d. This rate was the percentage of unique primer against overall unique primer

e. This rate was the percentage of unique primer with polymorphisms greater than or equal to 0.5 against unique primer

f. There were 7,406,619 primers in total, and the same primer might be divided into different regions and double counting due to the alternative splicing occurring in cucumber genome



PIC

Fig. 2 PIC value distribution of InDels in different genomic regions



Fig. 3 Size distribution of InDels in different genomic regions



Fig. 4 Size distribution of InDels in different genomic regions

highly polymorphic InDels with PIC  $\geq$  0.5 were 86,301. Of these, alleles  $\geq$  3 bp and  $\geq$  8 bp were 22,442 and 8,089, respectively (Supplementary Table 4). These InDels were easily genotyped using polyacrylamide and agarose gel electrophoresis. This detailed information on molecular markers, coupled with the convenience of Excel files, facilitates users in flexible marker selection, thereby improving efficiency in actual research applications.

#### GO analysis

A total of 1,823 genes with highly polymorphic InDel loci (PIC value  $\geq 0.5$ ) were identified in coding regions. Of these, 1,272 genes were assigned to 472 GO terms, with 136 significant terms. Among the terms, 47% belonged to the biological processes, 43% consisted of molecular functions and 10% comprised cellular components. A total of 1,531 of the 1,823 genes with highly polymorphic InDels were found in promoter regions. Of these, 1,012 genes were linked to 474 GO terms, with 113 significant terms. Of these GO terms, 64% belonged to the biological processes, 39% were involved in molecular functions, and 41% comprised cellular components. There were 264 overlapping genes between the coding and promoter regions, involving 96 GO terms, involved in metabolic and cellular processes. Furthermore, differences in GO terms between these regions were observed in DNA

metabolism, stress response, ion transport, transmembrane transport, and hydrolase activity (Fig. 5).

#### The experimental verification of indel markers

A total of 50 InDel markers with a PIC  $\geq$  0.12 (the minimum value in Supplementary Fig. 1) and a major allelic difference  $\geq$  3 bp were randomly selected for polymorphism verification in the DNA of 19 re-sequenced samples and 20 new lines, respectively (Fig. 6, Supplementary Fig. 1). The PIC of the InDel markers in 103 genome samples ranged from 0.12 to 0.88, averaging 0.53, while the number of allelic InDels ranged from 2 to 15, averaging 4.0. The PIC of the InDel markers in 19 re-sequenced DNA samples ranged from 0.12 to 0.64, with a mean of 0.47, while the number of allelic InDels ranged from 2 to 4, with a mean of 2.2. The PIC of the InDel markers in the DNA of 20 new lines ranged from 0.10 to 0.55, with a mean of 0.41, while the number of allelic InDels ranged from 2 to 3, with a mean of 2.0.

#### Discuss

InDels have received increasing attention in crops in recent years. However, the analysis of their variations is not as comprehensive as for SNPs. Their application as an ideal source for developing length polymorphism molecular markers has also been underutilized. In this study, we used cucumber as an example to construct a





Fig. 6 InDel polymorphisms experimental validation. Experimental validation of the InDel markers at position 9,831,160 on chromosome 1 and position 26,838,060 on chromosome 5 (PCR products from line 1 to line 20 originated from 19 cucumber inbred lines and one cucumber wild materials. Names and order are consistent with the previous description in Materials and Methods. M: marker DL2000)

genome-wide InDel variation map, then utilized these variations to develop and validate ideal InDel markers. This study maybe enhance the efficiency of research on cucumber populations and also provide a reference for InDel variation analysis and marker development in other crops, especially economic crops such as vegetables and fruits.

#### InDel hotspots of cucumber genome

Variation is the molecular basis of genetic and functional diversity. Identifying regions with concentrated variations can help to identify genetic resources and improve the efficiency of cucumber breeding. For instance, the genetic diversity of highly variable regions can be utilized to enhance heterosis, while in low-variation regions it can be increased to overcome the bottleneck effect in breeding germplasm resources. In our study, the identification of InDel variation hotspot regions revealed that regions with rich genetic diversity in the cucumber genome are relatively few, accounting for about 4.21% of the genome, which is significantly lower than that in the maize and rice [15, 16]. These hotspot regions are likely concentrated areas of important functional genes. Significant InDel variation hotspots have also been found in the

cucumber genome, including many important genes controlling key traits, such as CsY1 [37] and CsaARC5 [38] genes related to fruit peel colour, CsFT [39] and CsCRC [40] which control fruit quality, and *dm* [41] and *FW2.2* [42] genes which control disease resistance. Though the number of InDel and SNP hotspots was similar, 38 InDelonly hotspot regions did not overlap with SNP hotspot regions. The discovery of these regions provides new insights into solving the problem of inaccurate genetic diversity evaluation caused by the limitations of markers. This also indicates that when evaluating genetic diversity, the molecular basis of variation sources should be fully considered. This is because a single type of molecular marker may not completely reflect the molecular basis of genetic diversity and could potentially lead to an incomplete evaluation of germplasm resources, ultimately affecting their utilization.

#### Development of ideal indel markers in cucumber

For a long time, molecular markers have been powerful tools for studying genetic variation and key trait genes in cucumbers. Therefore, developing ideal InDel molecular markers, characterized by high density, high polymorphism and ease of detection, is necessary and offers new tools for cucumber breeding and strategies for molecular marker development in other vegetables. In this experiment, we developed 22,442 InDel primers, greatly surpassing the 134 InDel primers developed in 2013 [43]. The density of InDel markers in this study is higher than that reported for tomato [44] and cowpea [45]. High polymorphism increases the likelihood of detecting genetic differences and reduces primer synthesis costs. In our study, 18.71% of the InDels were polymorphic with the highest polymorphism rate of 0.96%, which is lower than maize [15] and rice [16], likely due to cucumber's simpler genetic background. The results indicate that it is necessary to use this method to identify highly polymorphic markers in species with low genetic diversity, such as cucumbers. Appropriate differences and ratios in amplification lengths can also improve the detection efficiency of agarose gel electrophoresis. In theory, under the condition of unchanged difference and gel concentration, the shorter the PCR amplification length, the higher the identification efficiency. Compared with SSR, InDels do not require primers to cross over repetitive sequences, so shorter PCR amplification lengths can be designed. Therefore, when designing primers for indel marker sites, we choose primers with the shortest amplification fragment under the premise of meeting the primer design conditions. The set of primers that were rigorously selected in this study are efficient length polymorphism primers.

The advantages of developing indel markers using contigs

Generally, in the development of InDel markers, sequence length affects the efficiency of InDel identification. Theoretically, the longer the fragment, the higher the probability of identifying larger InDel sites. For a given species, the number of available chromosomelevel genomes is often very limited. Even if the number of available reference genomes increases, it is still far less than the available small fragment data. For example, although there are hundreds of rice reference genomes, there are still tens of thousands of available small fragment data. In this study, most re-sequencing data fragment lengths are 90 bp. Such smaller fragment sizes limit the identification of larger InDel sites. Therefore, using contig data can effectively compensate for the limitations of small fragment sequencing data in InDel site identification. Moreover, during the contig assembly process, sequencing errors can be minimized to the greatest extent, improving the accuracy and completeness of sequencing data.

## Development of indel markers located within the gene in cucumber

The functional InDels in the genome can affect gene expression, thereby influencing biological traits and also play an important role in gene, biological breeding, and medical research. InDels located within the genome are more advantageous for genetic diversity assessment and marker-assisted selection (MAS). When assessing genetic effects, InDel markers derived within the gene are advantageous compared to markers derived near the gene. In MAS, 'within-gene markers' help prevent recombination that could prevent the separation of markers from target genes. Many key genes contain InDel variations, such as the Gn1a/OsCKX2 [46], GW2 [47], Chalk5 [48] and DEP1 [49] in rice, as well as Dwarf8 [50], qHO6/ DGAT1-2 [51] and Yr36/WKS1 [52] in maize and wheat, respectively. In this study, nearly half of the InDel markers were shown to be distributed within the gene region, with 6,380 highly polymorphic InDels found in the promoter regions and 899 in the coding regions. Interestingly, in coding regions, we found that 43% of the InDels are 3n-bp in length. Therefore, special attention should be paid to non-multiple-of-3 InDel sites in the research, as they are more likely to result in functional differences. Some important shape or functional gene of cucumber are recorded in Table 3 [53–70].

#### Conclusions

Here, we conducted a comprehensive identification of InDels within the cucumber genome and constructed the initial cucumber indel variation map. Based on this, we developed easy-to-use InDel markers that are distributed throughout the genome, with high polymorphism.

Table 3 Important shape and functional gene in cucumber

Gene	Candidate gene	Category	Location	References
fgr (Fragrance)	CsGy1G001790 (CsBADH)	Fragrance	Chr1: 1,161,602 0.1,166,880(+)	[53]
mf (Mango fruit)	CsGy1G007020 (CsWOX1)	Size and shape	Chr1: 4,487,766. 4,488,997(-)	[54]
cn (Carpel number)	CsGy1G014910 (CsCLV3)	Size and shape	Chr1: 10,815,721. 10,816,766(+)	[55]
m (Andromonoecious)	CsGy1G027100 (CsACS2)	Sex expression	Chr1: 25,520,318. 25,522,544(+)	[56]
sf2 (Short fruit 2)	CsGy2G010390 (CsSF2)	Size and shape	Chr2: 10,113,710. 10,123,269(+)	[57]
A (Androecious)	CsGy2G018140(CsACS11)	Sex expression	Chr2: 27,953,854. 27,957,003(+)	[58]
gl1 (Glabrous1)	CsGy3G031820 (CsGL1)	Trichome	Chr3: 31,243,052. 31,245,810(-)	[59]
w (White skin color)	CsGy3G044470 (CsAPRR2)	Epidermal feature	Chr3: 41,171,141. 41,178,296(-)	[60]
B (Black spine)	CsGy4G001040 (CsMYB60)	Epidermal feature	Chr4: 635,198. 635,650(-)	[61, 62]
si (short internode)	CsGy4G022710 (CsVFB1)	Architecture	Chr4: 29,078,312. 29,079,919(-)	[63]
dm1 (Resistance to Pseudoperonospora cubensis)	CsGy5G003280 (CsSGR)	Oomycete resistance	Chr5: 2,147,948. 2,149,557(+)	[64]
bt (bitter fruit)	CsGy5G003340 (CsBt)	Bitterness	Chr5:2,198,470 0.2,199,474 (-)	[13]
Ts1 (Tubercle size)	CsGy5G017890 (CsTbs1)	Epidermal feature	Chr5: 24,119,025. 24,119,573(+)	[57]
Tu (Warty)	CsGy5G019590 (CsTu)	Epidermal feature	Chr5: 25,945,387. 25,946,028(+)	[65]
pm5.1 (Resistance to Podosphaera fusca)	CsGy5G026660 (CsMLO1)	Fungal resistance	Chr5: 30,523,505. 30,528,482(-)	[66]
bi (bitterfree)	CsGy6G007190 (CsBi)	Bitterness	Chr6: 6,144,810 0.6,150,753 (-)	[67]
F (Femaleness)	CsGy6G028780 (CsACS1G)	Sex expression	Chr6: 27,576,407. 27,592,191(-)	[68]
scl-1 (Small and cordate leaf)	CsGy7G005090 (CsSCL1)	Leaf	Chr7: 3,794,067. 3,798,566(+)	[69, 70]

We believe that these molecular markers will have high application value in the future, especially in disease resistance breeding and quality breeding and can provide simpler and more effective tools for studying genetic effects. Indel markers are the most densely distributed among length polymorphism markers; however, their overall density remains lower than that of SNP markers. Additionally, the efficiency of screening large-scale indel marker loci is relatively low. Future research should place greater emphasis on improving this aspect.

#### **Supplementary Information**

The online version contains supplementary material available at https://doi.or g/10.1186/s12864-025-11584-z.

Supplementary Material 1 Supplementary Table 1: Classification of InDels from 115 re-sequenced cucumber samples

Supplementary Material 2 Supplementary Table 2: Identified of InDel and SNP hotspots of cucumber genome

Supplementary Material 3 Supplementary Table 3: The next-generation sequencing date and contig assembly of cucumber samples

Supplementary Material 4 Supplementary Table 4: Information of InDel markers (cucumber v2/v3/v4)

Supplementary Material 5 Supplementary File 1: Primer design selection via e-PCR alignment

Supplementary Material 6 Supplementary Figure 1: Experimental validation of polymorphism of 48 InDel

#### Acknowledgements

We highly appreciate Prof. Sanwen Huang's help for providing us with DNA samples, and we also thank Professors Aimin Wei, Yike Han, and Xinghua Cui from the Tianjin Kernel Cucumber Research Institute for their guidance and technical support in the experiments.

#### Author contributions

JC Y, PX M, and H M completed the analysis and summary of the data, as well as the writing of the manuscript, and were major contributors in writing the manuscript. XY W, J Y, SH F, WZ G, and RF B were responsible for constructing the article's framework and providing data during the calculation process. WJ D, HM W, ZQ L, and SJ Y improved the quality of the manuscript. All authors read and approved the final manuscript.

#### Funding

This work was supported by National Key Research and Development Program of China (2022YFD1601501), Major Science and Technology Program in Sichuan Province (2023YFN0006), Natural Science Foundation Project of Sichuan Provincial (2022NSFSC0151), the Major Science and Technology Project of Sichuan Province (2022ZDZX0013), and National Key Research and Development Program of China (2023YFD1201100).

#### Data availability

The datasets underlying this article are available in National Center for Biotechnology Information (NCBI), at http://www.ncbi.nlm.nih.gov/sra/?term= SRA056480, and in Cucumber Multi-omics Database (Cucumber-DB), at http:// www.cucumberdb.com.

#### Declarations

**Ethics approval and consent to participate** Not applicable.

#### **Consent for publication** Not applicable.

Competing interests

The authors declare no competing interests.

#### Received: 29 November 2024 / Accepted: 8 April 2025 Published online: 18 April 2025

#### References

- Lv J, Qi JJ, Shi QX, Shen D, Zhang SP, Shao GJ, et al. Genetic diversity and population structure of cucumber (Cucumis sativus L). PLoS ONE. 2012;7:e46919.
- 2. Pan Y, Qu SP, Bo KL, Gao ML, Haider KR, Weng YQ, et al. QTL mapping of domestication and diversifying selection related traits in round-fruited

semi-wild Xishuangbanna cucumber (Cucumis sativus L. Var. xishuangbannanesis). Theor Appl Genet. 2017;130:1531–48.

- Li XX, Zhu DW, Du YC, Shen D, Kong QS, Song JP, et al. Studies on genetic diversity and phylogenetic relationship of cucumber (Cucumis sativus L.) germplasm by AFLP technique. Acta Horticulturae Sinica. 2004;31:309–14. (In Chinese).
- Wang HZ, Li SJ, Liu XF, Li P, Huo ZR, Guan W. AFLP markers of cucumber anthracnose resistance-related gene. Acta Horticulturae Sinica. 2007. (In Chinese).
- Lu BR, Cai XX, Jin X. Efficient indica and Japonica rice identification based on the indel molecular method: its implication in rice breeding and evolutionary research. Prog Nat Sci. 2009;19:1241–52.
- Li XX, Zhu DW, Du YC, Zhang GP, Shen D. Genetic diversity and phylogenetic relationship of cucumber (Cucumis sativus L.) germplasm based on RAPD analysis. J Plant Genetic Resour. 2004;5:147–52. (In Chinese).
- Zhang HX, Zhang HY, Yu GJ, Zhang F, Mao AJ, Wang YJ, et al. Identification RAPD markers linked to fusarium wilt resistance Gne in cucumber. Acta Agriculturae Boreali-Sinica. 2006;21:121–3. (In Chinese).
- Cavagnaro PF, Senalik DA, Yang LM, Simon PW, Harkins TT, Kodira CD, et al. Genome-wide characterization of simple sequence repeats in cucumber (Cucumis sativus L). BMC Genomics. 2010;11:569.
- Yong JP, Li YH, Meng YJ, Zhong YH, Cheng ZH, Cheng P. The simple sequence repeat (SSR) and sequence-tagged sites (STS) markers linked to the compact gene (cp) in cucumber (Cucumis sativus L). J Agricultural Biotechnol. 2013;22:1152–58. (In Chinese).
- Yundaeng C, Somta P, Tangphatsornruang S, Chankaew S, Srinives P. A single base substitution in BADH/AMADH is responsible for fragrance in cucumber (Cucumis sativus L.), and development of SNAP markers for the fragrance. Theor Appl Genet. 2015;128:1881–92.
- Zhang HM, Qu W, Jing HJ, Ding XT, Yu JZ. Genetic diversity analysis of 23 cucumber germplasms and screening of core germplasm resources using indel markers. Acta Agriculturae Shanghai. 2019;35:28–33. (In Chinese).
- Lu X, Liu MH, Deng ZJ, Sun QY, Xia YC, Li WH, et al. Genetic diversity analysis of cucumber germplasm resources based on indel markers. Jiangsu Agricultural Sci. 2021;49:49–54. (In Chinese).
- Zhang SP, Miao H, Cheng ZR, Zhang ZH, Wu J, Sun RF, et al. The Insertiondeletion (Indel) marker linked to the fruit bitterness gene (*Bt*) in cucumber. J Agricultural Biotechnol. 2011;19:649–53. (In Chinese).
- Nie JT, Li XD, Yao YK, Pan SJ, He HL, Liu SH et al. InDel markers for identification of cucumber powdery mildew resistance. China Vegetables. 2015: 26–30. (In Chinese).
- Liu J, Qu JT, Yang C, Tang DG, Li JW, Lan H, et al. Development of genomewide insertion and deletion markers for maize, based on next-generation sequencing data. BMC Genomics. 2015;16:601.
- Liu J, Li JW, Qu JT, Yan SY. Development of Genome-Wide insertion and deletion polymorphism markers from Next-Generation sequencing data in rice. Rice (N Y). 2015; 8: 63.
- Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, et al. Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved LncRNAs associated with agriculture traits. Plant J. 2015;84:404–16.
- Zhang L, Fu MM, Li WY, Dong YB, Zhou Q, Wang QL, et al. Genetic variation in *ZmKW1* contributes to kernel weight and size in Dent corn and popcorn. Plant Biotechnol J. 2024;22:1453–67.
- 19. Zhao D, Zheng HW, L JJ, Wan MY, Shu K, Wang WH, et al. Natural variation in the promoter of *GmSPL9d* affects branch number in soybean. Int J Mol Sci. 2024;25:5991.
- Wang N, Liu WJ, Mei ZX, Zhang SH, Zou Q, Yu L, et al. A functional indel in the WRKY10 promoter controls the degree of flesh red pigmentation in Apple. Adv Sci (Weinh). 2024;11:e2400998.
- Fang T, Wang MZ, He RJ, Chen QW, He DY, Chen XR, et al. A 224-bp indel in the promoter of *PeMYB115* accounts for anthocyanin accumulation of skin in passion fruit (Passiflora spp). J Agric Food Chem. 2024;72:10138–48.
- Zhu YF, Zhu GT, Xu R, Jiao ZX, Yang JW, Lin T, et al. A natural promoter variation of SIBBX31 confers enhanced cold tolerance during tomato domestication. Plant Biotechnol J. 2023;21:1033–43.
- Xi'ou X, Cao BH, Li GN, Lei JJ, Chen QH, Jiang J, et al. Functional characterization of a putative bacterial wilt resistance gene (*RE-bw*) in eggplant. Plant Mol Biology Report. 2015;33:1058–73.
- 24. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. 2011;21:830–9.

- 25. Ding S, Wang SP, He K, Jiang MX, Li F. Large-scale analysis reveals that the genome features of simple sequence repeats are generally conserved at the family level in insects. BMC Genomics. 2017;18:848.
- Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJ. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics. 1991;10:654–60.
- Qu JT, Liu J. A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data. BMC Genomics. 2013;6:403.
- Liu J, Qu JT, Hu K, Zhang L, Li JW, Wu B, et al. Development of genomewide simple sequence repeat fingerprints and highly polymorphic markers in cucumbers based on next-generation sequence data. Plant Breeding. 2015;134:605–11.
- 29. Guan JT, Miao H, Zhang ZH, Dong SY, Zhou Q, Liu XP, et al. A near-complete cucumber reference genome assembly and Cucumber-DB, a multi-omics database. Mol Plant. 2024;17:1178–82.
- Qi JJ, Liu X, Shen D, Miao H, Xie BY, Li XX, et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet. 2013;45:1510–5.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100:9440–5.
- Yoav B, Yosef H. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol). 1995;57:289–300.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19:1117–23.
- Liu J, Qu JT, Hu K, Zhang L, Li J, Wu B, et al. Development of genomewide simple sequence repeat fingerprints and highly polymorphic markers in cucumbers based on next-generation sequence data. Plant Breeding. 2015;134:605–11.
- 35. Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME. Optimizing parental selection for genetic linkage maps. Genome. 1993;36:181–6.
- Du Z, Zhou X, Ling Y, Zhang ZH, Su Z. AgriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 2010;38:W64–70.
- Han YK, Zhao FY, Gao S, Wang XY, Wei AM, Chen ZW, et al. Fine mapping of a male sterility gene ms-3 in a novel cucumber (Cucumis sativus L.) mutant. Theor Appl Genet. 2018;131:449–60.
- Zhou Q, Wang SH, Hu BW, Chen HM, Zhang ZH, Huang SW. An ACCUMULA-TION AND REPLICATION OF CHLOROPLASTS 5 gene mutation confers light green Peel in cucumber. J Integr Plant Biol. 2015;57:936–42.
- Wen CL, Zhao WS, Liu WL, Yang LM, Wang YH, Liu XW, et al. CsTFL1 inhibits determinate growth and terminal flower formation through interaction with CsNOT2a in cucumber. Development. 2019;146:1–12.
- Che G, Pan YP, Liu XF, Li M, Zhao JY, Yan SS, et al. Natural variation in CRABS CLAW contributes to fruit length divergence in cucumber. Plant Cell. 2023;35:738–55.
- 41. Li LX, He HQ, Zou ZR, Li YH. QTL analysis for downy mildew resistance in cucumber inbred line PI 197088. Plant Dis. 2018;102:1240–5.
- 42. Dong JP, Xu J, Xu XW, Chen XH. Inheritance and quantitative trait locus mapping of fusarium wilt resistance in cucumber. Front Plant Sci. 2019;10:1425.
- 43. Li SG, Shen D, Liu B, Qiu Y, Zhang ZH, Wang HP, et al. Development and application of cucumber indel markers based on genome re-sequencing. J Plant Genetic Resour. 2013;14:278–83. (In Chinese).
- Zhang GR, Tang YP, Yang T, Pati GL, Wang BK, Li N, et al. Development of indel markers for tomato based on Re-sequencing data. Mol Plant Breed. 2019;17:4692–7. (In Chinese).
- 45. Yang Y, Li TY, Li GJ, Chen HC, Shen Z, Shen Z, et al. Development and application of Insertion-Deletion (InDel)Markers in asparagus bean based on whole genome Re-sequencing data. Acta Horticulturae Sinica. 2022;49:778–90. (In Chinese).
- Li YB, Fan CC, Xing YZ, Yun P, Luo LJ, Yan B, et al. Chalk5 encodes a vacuolar H (+)-translocating pyrophosphatase influencing grain chalkiness in rice. Nat Genet. 2014;46:398–404.
- Ashikari M, Sakakibara H, Lin SY, Yamamoto T, Takashi T, Nishimura A, et al. Cytokinin oxidase regulates rice grain production. Science. 2005;309:741–5.
- Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet. 2007;39:623–30.
- 49. Huang XZ, Qian Q, Liu ZB, Sun HY, He SY, Luo D, et al. Natural variation at the DEP1 locus enhances grain yield in rice. Nat Genet. 2009;41:494–7.

- Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann MF, Barraud S, et al. Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics. 2006;172:2449–63.
- Hao XM, Li XW, Yang XH, Li JS. Transferring a major QTL for oil content using marker-assisted backcrossing into an elite hybrid to increase the oil content in maize. Mol Breeding. 2014;34:739–48.
- Fu DL, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen XM, et al. A kinase-START gene confers temperature-dependent resistance to wheat Stripe rust. Science. 2009;323:1357–60.
- Chutintorn Y, Somta P, Tangphatsornruangg S, Chankaes S, Srinives P. A single base substitution in BADH/AMADH is responsible for fragrance in cucumber (Cucumis sativus L.), and development of SNAP markers for the fragrance. Theor Appl Genet. 2015;128:1881–92.
- Niu HH, Liu XF, Tong C, Wang H, Li S, Lu L, et al. The WUSCHEL-related homeobox1 gene of cucumber regulates reproductive organ development. J Exp Bot. 2018;69:5373–87.
- Che G, Gu R, Zhao JY, Liu XF, Song XF, Zi ZH, et al. Gene regulatory network controlling carpel number variation in cucumber. Development. 2020;147:184788–801.
- Li Z, Wang S, Tao QY, Pan JS, Si LT, Gong ZH, et al. A putative positive feedback regulation mechanism in CsACS2 expression suggests a modified model for sex determination in cucumber (Cucumis sativus L). J Exp Bot. 2012;63:4475–84.
- Wang YH, Bo KL, Gu XF, Pan JS, Li YH, Chen JF, et al. Molecularly tagged genes and quantitative trait loci in cucumber with recommendations for QTL nomenclature. Hortic Res. 2020;1:3–23.
- Boualem A, Troadec C, Camps C, Lemhemdi A, Morin H, Sari MA, et al. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. Science. 2015;350:688–91.
- Li Q, Cao CX, Zhang CJ, Zheng SS, Wang ZH, Wang L, et al. The identification of cucumis sativus glabrous 1 (*CsGL1*) required for the formation of trichomes uncovers a novel function for the homeodomain-leucine zipper I gene. J Exp Bot. 2015;66:2515–26.
- 60. Yao XH, Li HJ, Nie J, Liu H, Guo YC, Lv LJ, et al. Disruption of the amino acid transporter *CsAAP2* inhibits auxin-mediated root development in cucumber. New Phytol. 2023;239:639–59.

- 61. Liu MY, Zhang CJ, Duan LX, Luan QQ, Li JL, Yang A, et al. CsMYB60 is a key
- regulator of flavonols and proanthocyanidans that determine the colour of fruit spines in cucumber. J Exp Bot. 2019;70:69–84.
- 62. Li JL, Luan QQ, Han J, Zhang CJ, Liu MY, Ren ZH. *CsMYB60* directly and indirectly activates structural genes to promote the biosynthesis of flavonols and proanthocyanidins in cucumber. Hortic Res. 2020;1:103–18.
- 63. Liu XF, Chen JC, Zhang XL. Genetic regulation of shoot architecture in cucumber. Hortic Res. 2021;8:143–57.
- 64. Pan JS, Tan JY, Wang YH, Zheng XY, Owens K, Li D, et al. STAYGREEN (*CsSGR*) is a candidate for the anthracnose (Colletotrichum orbiculare) resistance locus Cla in *Gy14* cucumber. Theor Appl Genet. 2018;131:1577–87.
- 65. Yang S, Wen CL, Liu B, Cai YL, Xue SD, Bartholomew ES, et al. A *CsTu-TS1* regulatory module promotes fruit tubercule formation in cucumber. Plant Biotechnol J. 2019;17:289–301.
- Nie JT, Wang YL, He HL, Guo CL, Zhu WY, Pan J, et al. Loss-of-Function mutations in *CsMLO1* confer durable powdery mildew resistance in cucumber (Cucumis sativus L). Front Plant Sci. 2015;22:1155–70.
- Anwar S, Siddique R, Ahmad S, Haider MZ, Ali H, Sami A, et al. Genome wide identification and characterization of *Bax inhibitor-1* gene family in cucumber (Cucumis sativus) under biotic and abiotic stress. BMC Genomics. 2024;25:1032–50.
- Knopf RR, Trebitsh T. The female-specific Cs-ACS1G gene of cucumber. A case of gene duplication and recombination between the non-sex-specific 1-aminocyclopropane-1-carboxylate synthase gene and a branched-chain amino acid transaminase gene. Plant Cell Physiol. 2006;47:1217–28.
- Wang CH, Li J, Fang K, Yao HX, Chai XW, Du YL, et al. CsHLS1-CsSCL28 module regulates compact plant architecture in cucumber. Plant Biotechnol J. 2024;22:1724–39.
- Vielba JM, Carmen DS, Ferro E, Rico S, Lamprecht M, Abarca D, et al. *7*is differentially regulated upon maturation in chestnut microshoots and is specifically expressed in rooting-competent cells. Tree Physiol. 2011;31:1152–60.

#### Publisher's note

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