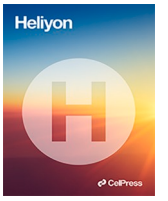
Heliyon 10 (2024) e29521

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

# Genetic diversity, population structure and marker-trait associations in Indian kale (*Brassica oleracea* L. gp. acephala) using cross-species microsatellite markers

Geetika Malik [a,*], Asma Jabeen [a], Javid Iqbal Mir [a], Rafiq Ahmad Shah [a], Mohd Abas Shah [a], Vishal Dinkar [a], Muneer Ahmad Sheikh [a], Ravinder Kumar [b,**], Om Chand Sharma [a], Mahendra Kumar Verma [a]

[a] *Indian Council of Agricultural Research-Central Institute of Temperate Horticulture, Srinagar, Jammu & Kashmir, India, 191132*
[b] *Indian Council of Agricultural Research-Indian Agricultural Research Institute, New Delhi, India, 110012*

ARTICLE INFO

*Keywords:*
Kale
Genetic diversity
Marker-trait association
Population structure
Linkage disequilibrium
SSR markers

ABSTRACT

Kale is known for its exceptional nourishing and functional benefits to human body. However, it is an understudied species from genomic as well as agronomic aspects. It is important to characterize niche kale germplasms around the world to systematically conserve and utilize its genetic variability, especially for commercial traits in the interest of growers, consumers and industry. With this view, genomic and phenotypic characterizations of 62 Kashmiri kale accessions including popular landraces were done to estimate and partition genetic diversity, understand trait relationships, develop population structure and divulge marker-trait associations of economic significance. Sixty-six cross species microsatellite (SSR) markers within *Brassica* genus amplified 269 alleles in the germplasm. Their polymorphic information content (PIC) ranged from 0.00078 to 0.953 with an average of 0.407. The population structure analysis and neighbour joining tree clustering categorized the germplasm into three sub-populations. AMOVA revealed more within-population variance (67.73 %) than among-populations (32.27 %) variance. The principal component analysis (PCA) involving 24 agronomical traits revealed seven PCs (PC1 to PC7) having Eigen values more than 1, which explained a cumulative variation of 69.21 %. Association mapping with respect to these 24 agronomical traits using mixed linear model and general linear model revealed six overlapping significant marker-trait relationships with five being significant at probability value of 0.001/0.0001. The highly significant associations of two SSRs with economically important traits (siliqua length and seed weight) significantly correlated/related with leaf yield and seed yield were revealed for their possible utilization in marker assisted breeding for higher leaf and seed yields.

## 1. Introduction

Referred to as the dog of the plant world, *Brassica oleracea* L. *oleracea* is the most diverse species in plant kingdom just as dog is the

* Corresponding author.
** Corresponding author.
*E-mail addresses:* geetika.malik@icar.gov.in (G. Malik), chauhanravinder97@gmail.com (R. Kumar).

https://doi.org/10.1016/j.heliyon.2024.e29521
Received 21 December 2023; Received in revised form 8 April 2024; Accepted 9 April 2024
Available online 18 April 2024

most diverse animal species [1]. Native to South and West Europe and thriving on sea cliffs, *Brassica oleracea* L. *oleracea* is commonly called wild cabbage and is the progenitor of diverse morphotypes that rule the world-wide vegetable fields and gardens today. It has been domesticated into various cultivars, the best known among which are cauliflower (Botrytis group), cabbage (Capitata group), broccoli (Italica group), Brussels' sprouts (Gemmifera group), kohlrabi (Gongyloides group) and kale & collards (Acephala group). Also known as colworts by the 19th century English, the Acephala (without head) group also became favourite of English after the

**Table 1**
Germplasm of kale (*Brassica oleracea* L. Gp. Acephala) used in the study.

| Accession name | Number assigned | Origin | Identity |
|---|---|---|---|
| CITH-KC-1/17 | 1 | Kashmir | IC-0650676 |
| CITH-KC-1/21 | 2 | Kashmir | IC-0650674 |
| CITH-KC-3 | 3 | Kashmir | IC-0650702 |
| CITH-KC-9 | 4 | Kashmir | IC-0650678 |
| CITH-KC-14 | 5 | Kashmir | IC-0650680 |
| CITH-KC-18 | 6 | Kashmir | IC-594205 |
| CITH-KC-21 | 7 | Kashmir | Collection |
| CITH-KC-26 | 8 | Kashmir | IC-0650694 |
| CITH-KC-26R | 9 | Kashmir | Selection |
| CITH-KC-40 | 10 | Kashmir | IC 594203 |
| CITH-KC-42 | 11 | Kashmir | IC-0650701 |
| CITH-KC-44 | 12 | Kashmir | IC 594207 |
| CITH-KC-48 | 13 | Kashmir | IC-0650683 |
| NW-Sag-36R | 14 | Kashmir | IC-0650681 |
| NW-Sag-21 | 15 | Kashmir | IC-0650686 |
| *Kawdari* | 16 | Kashmir | Landrace (IC-0650671) |
| NW-Sag-27 | 17 | Kashmir | IC-0650670 |
| CITH-KC-24KK | 18 | Kashmir | Selection |
| CITH-KC-38 | 19 | Kashmir | IC-0650673 |
| NW-Sag-41 | 20 | Kashmir | IC-0650669 |
| Pusa-Sag-1 | 21 | Himachal Pradesh | Collection |
| NW-Sag-1 | 22 | Kashmir | IC-0650675 |
| Pusa-Sag-2 | 23 | Kashmir | Selection |
| NW-Sag-29 | 24 | Kashmir | IC-0650713 |
| HW-1 | 25 | Kashmir | IC-0650677 |
| Pusa-Sag-3 | 26 | Kashmir | Collection |
| *GM Dari* | 27 | Kashmir | IC-0650682 (Landrace) |
| CITH-KC-Sel-3 | 28 | Kashmir | IC-0650711 |
| *Khanyari* (dwarf) | 29 | Kashmir | Selection |
| NW-Sag-39 | 30 | Kashmir | IC-0650687 |
| CITH-KC-4/21 | 31 | Kashmir | IC-0650672 |
| NW-Sag-38 | 32 | Kashmir | IC-0650685 |
| NW-Sag-44 | 33 | Kashmir | IC-0650689 |
| NW-Sag-40 | 34 | Kashmir | IC-0650688 |
| NW-Sag-24 | 35 | Kashmir | IC-0650684 |
| NW-Sag-15 | 36 | Kashmir | IC-0650692 |
| *Hanz haaq* | 37 | Landrace | IC-0650693 (Landrace) |
| NW-Sag-30 | 38 | Kashmir | IC-0650691 |
| Siberian | 39 | Exotic | Collection from private farm |
| CITH-KC-2 | 40 | Kashmir | IC-0650695 |
| *Khanyari* | 41 | Kashmir | IC-0650697 (Landrace) |
| CITH-KC-24 | 42 | Kashmir | IC 594206 |
| CITH-KC-Sel-5 | 43 | Kashmir | IC-0650699 |
| NW-Sag-33 | 44 | Kashmir | IC-0650700 |
| NW-Sag-23 | 45 | Kashmir | IC-0650698 |
| CITH-KC-4-Sel1/15 | 46 | Kashmir | Selection |
| CITH-KC-16 | 47 | Kashmir | IC-0650710 |
| NW-Sag-42 | 48 | Kashmir | IC-0650704 |
| NW-Sag-4 | 49 | Kashmir | IC-0650706 |
| CITH-KC-7 | 50 | Kashmir | IC 594208 |
| CITH-KC-Sel-1 | 51 | Kashmir | IC-0650705 |
| CITH-KC-20 | 52 | Kashmir | Collection |
| NW-Sag-20R | 53 | Kashmir | IC-0650679 |
| CITH-KC-17 | 54 | Kashmir | Collection |
| Japanese Green | 55 | Exotic | Collection from private farm |
| CITH-KC-8 kh(R) | 56 | Kashmir | Selection |
| CITH-KC-11 | 57 | Kashmir | Collection |
| CITH-KC-6/21 | 58 | Kashmir | IC-0650703 |
| HW-5 | 59 | Kashmir | IC-0650708 |
| CITH-KC-53 | 60 | Kashmir | IC-0650709 |
| Pusa-Sag-4 | 61 | Kashmir | Selection |
| CITH-KC-3/21 | 62 | Kashmir | IC-0650707 |

popularity of heading types (broccoli, cauliflower, cabbage) [2]. The present-day var. *viridis* of convar. *acephala* is known as collard and is the American derivative of English word 'colewort'. The variety *acephala* that too comes under convar. *acephala* is known as kale. Sometimes, the term kale is used collectively for both collard and kale. Kale (*Brassica oleracea* L. group Acephala), the king of leafy vegetables, with a century-old history, belongs to *Brassicaceae* family [3]. It is grown throughout the world with significant European areas, central and northern parts of Peninsula and parts of the Black Sea coast [4–7]. Owing to its nutraceutical and medicinal potential, kale has gained significant attention world-over and has been recognized as an exceptional source of phytochemicals. Kale, renowned as 'Superfood', with its health benefits and valuable nutritional composition, is a rich source of vitamin A, vitamin K, carotenoids, glucosinolates, lutein, zeaxanthin, iron, magnesium and protein [8–11]. It is loaded with powerful antioxidants like quercetin and kaempferol [12,13] and is a good source of folic acid, the pregnancy vitamin. Due to these properties, pharma/nutraceutical companies are now coming up with dietary supplement formulations prepared from kale [14]. Kale bread, kale juice, kale puree and kale microcapsules are some of the famous value-added products and formulations [15].

Nutraceutical and bioactive profiling is very important for crop improvement programs [16] and diversity assessment [17] of horticultural crops. Kale in India is primarily grown, traded and consumed in the Union Territory of Jammu and Kashmir, particularly in Kashmir region where it is a traditional crop. Probably, the most comprehensive known diversity of kale (*Brassica oleracea* gp. Acephala) in India is found in Kashmir valley embedded in Indian Himalayan Region. Kashmiri kale germplasm is morphologically highly diverse with several landraces markedly different from each other. Among many landraces, *Khanyari* is known for its succulent and puckered leaves, indeterminate columnar growth and adaptability to warmer temperatures, *GM Dari* bears large flat leaves, has rosette appearance and is highly adapted to cool season. *Hanz Haaq* is a short stature plant with rosette of tightly arranged leaves that keep appearing indefinitely ensuring multiple harvests. Since, kale is a highly cross-pollinated crop, it is common to find useful variants in farmers' fields. The accessions developed from these variant collections have also been found to have considerable phenotypic diversity. This diverse range of kale, whether among landraces or inbred accessions, can be effectively utilized for its genetic enhancement. However, it needs to be genetically characterized first.

Genetic characterization of any crop species germplasm is vital to utilizing desirable genes and developing their conservation strategies [18]. It helps carve population structures, elucidate genetic diversity, measure genetic relatedness and differences and understand evolution of populations and species on the whole. Nowadays, molecular markers play central role in achieving these goals. The most widely used molecular markers for the purpose are simple sequence repeats (SSR) or microsatellites, as they are co-dominant, repeatable, often polymorphic, and widely distributed throughout the genome [19].

Association mapping has become a routine practice in linking available markers with quantitative trait loci (QTLs) of importance in crop species [20]. Because association mapping has a higher mapping resolution than linkage mapping, it is frequently used to identify QTLs in crops [21]. Association mapping is based on linkage disequilibrium (LD) studies to evaluate statistically significant marker-trait associations (MTA). At the population level, association mapping takes advantage of past and evolutionary recombination. Other benefits include the ability to analyse several alleles at once and review numerous recombination events, in numerous generations, and in natural populations without creating a mapping population. Recently, MTA have been established using association mapping for quality and yield traits in crops such as tomato [22], maize [23], *Brassica napus* [20] watermelon [24], spinach [25] and *Brassica rapa* [26].

It has been hypothesized that landraces and naturally occurring variants in farmer fields are more diverse than modern bred cultivars and thus better sources of genetic variation for the purpose of improving a crop genetically [27,28]. Therefore, this study on Kashmiri kale landraces and inbred lines developed from farmer field collections was designed to estimate genetic diversity, attempt clustering of groups, develop population structure and partition the total variation into conveniently useable components for possible use in kale breeding and conservation. Secondly, the microsatellites or the SSRs are known to be highly conserved across species, more so within the same genus, and cross reproducible in species with no known SSRs. It was thus pertinent to study SSRs developed in other *Brassica* species for their reproducibility and practical utility in kale in the form of associations with economically important leaf and seed yield related traits besides other traits of interest.

## 2. Materials and methods

### 2.1. Germplasm

The germplasm has been developed from collections made from farmer fields across Kashmir valley and a few collected from other sources. Each accession is being maintained in active field gene bank at ICAR-Central Institute of Temperate Horticulture, Srinagar, Jammu and Kashmir in Karewa soils 33.5901°N and 74.4780 °E at 1500 m above mean sea level for more than 2–15 years. The accessions or germplasm lines are being inbred every year by caging 3 to 6 true-to-type plants together to circumvent cross-pollination with other lines. The details of the kale genotypes in presented in Table 1.

This germplasm was evaluated for 24 metric traits leaf blade length/LBL (cm), leaf blade diameter/LBD (cm), leaf blade number of incisions/LBNI, leaf blade thickness/LBT (mm), total number of leaves/NL, number of leaf lobes/NLO, petiole length/PeL (cm), petiole diameter/PeD (cm), vegetative stem length/VSL (cm), vegetative stem width/VSW (cm), leaf yield/LY (q/ha), initiation of flowering/IF (days), end of flowering/EF (days), flowering period/FP (days), size of floral buds/SFB, floral stalk length/FSL (cm), number of floral buds per plant/NFBPP, days to maturity/DM, siliqua length/SL (cm), siliqua width/SW (cm), number of seeds per siliqua/NSS and 1000 seed weight/1000 SW (g), plant height/PH (cm) and canopy diameter/CD (cm) and 26 non-metric traits: plant shape, plant stature & main stem branching, leaf blade shape, leaf blade margin, leaf color, leaf angle (degree), leaf apex shape, leaf midrib curvature, leaf blade curvature, leaf midrib colour, leaf anthocyanin colouration, leaf undulation, leaf blade curling, leaf pubescence, stem

colour, petiole colour, inflorescence axis, floral stalk branching, floral bud color, flower color, flower size, silique attitude, silique surface, silique colour before drying and seed colour. Some of these traits have been shown in Fig. 1.

### 2.2. Molecular characterization

#### 2.2.1. DNA extraction

The method described by Doyle and Doyle [29] was used to extract genomic DNA from 1g of young leaf tissue of different kale genotypes. The integrity and quality of genomic DNA was checked on 0.8 % agarose gel. Nano-drop spectrophotometer was also used to examine the amount and purity of DNA (ThermoScientific, Waltham, MA, USA). For subsequent use, high-quality DNA was diluted to a concentration of 25 ng/μl.

#### 2.2.2. SSR amplification

After literature survey, 75 SSR markers were selected based on the reported polymorphic information content (PIC) values. The PCR were carried out in a thermal cycler (Takara, Japan) using 35 cycles of initial denaturation at 94 °C for 30 s, annealing at 45–58 °C for 30 s, extension at 72 °C for 1 min, and final extension at 72 °C for 10 min. The PCR products were resolved in a 3 % agarose (MolBio

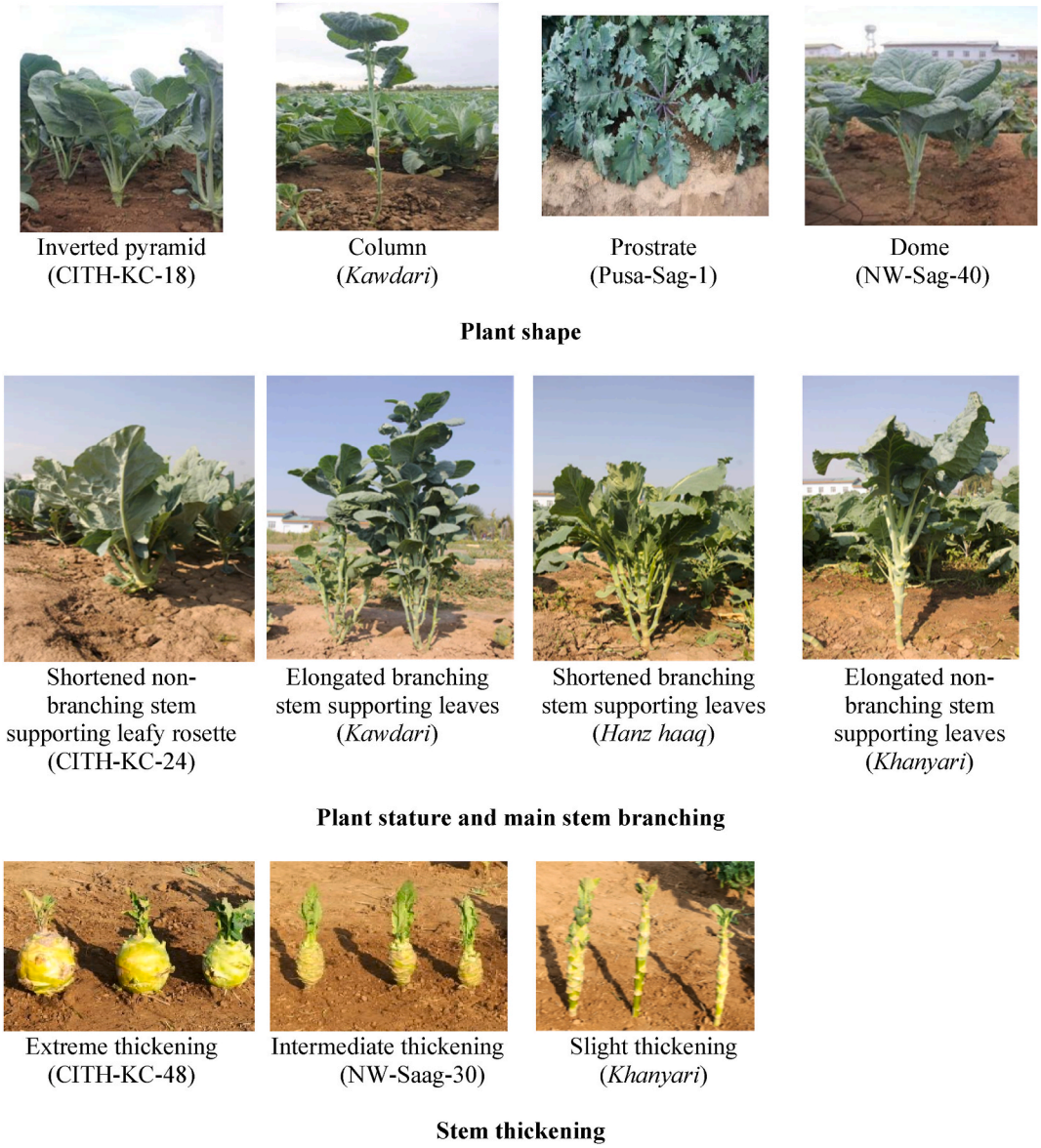



**Fig. 1.** Some non-metric traits expressed by the different kale accessions under study.

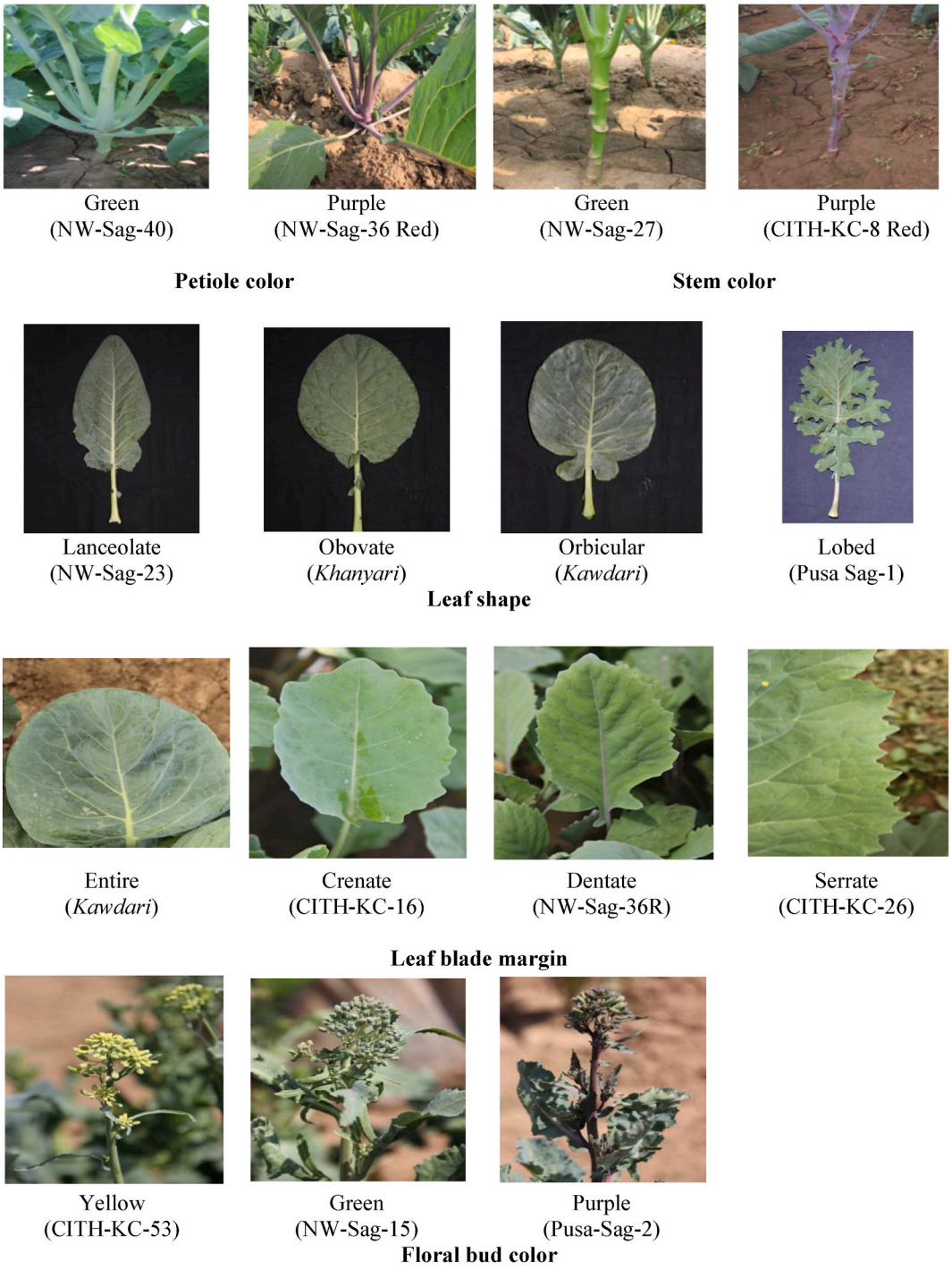
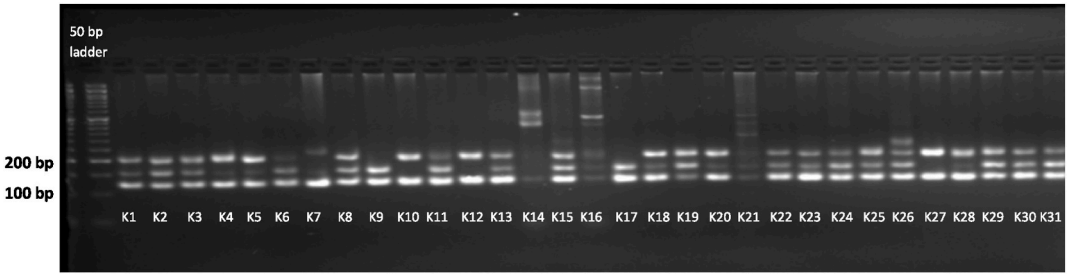



**Fig. 1.** (*continued*).

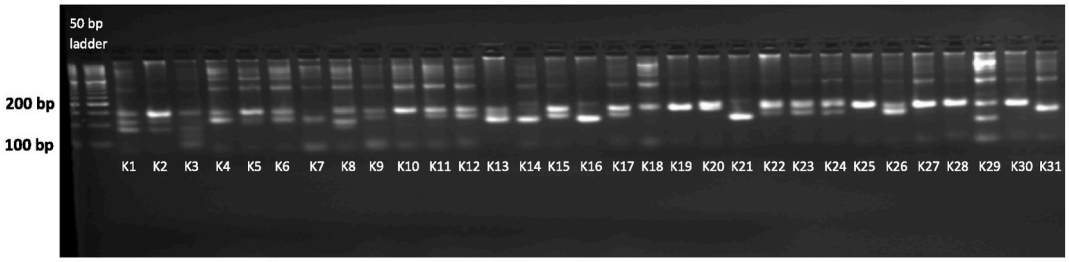
HiMedia) gel at constant 120 V for 2.5 h, and the size of the amplicon was determined using a 50 bp DNA ladder (HiMedia). The bands were examined under UV light using the gel documentation unit (Bio-Rad Laboratories Inc., USA). During the course of study, nine markers amplified none of the DNA samples. So, eventually only 66 markers contributed to the study (Table 2). These SSRs belonged to the chromosomes of *Brassica napus*, *B. oleracea*, *B. rapa*, *B. nigra* and *Raphanus sativus*. Fig. 2(a and b) display bands that were amplified using the primers Ol10-G08 and Ol12-G04.

**Table 2**
List of sixty-six microsatellite markers used in the study along with their chromosomal location, annealing temperature and primer sequence.

| SSR no. | Marker | Chromosomal location | AT | Forward primer (5′–3′) | Reverse primer (5′–3′) |
|---|---|---|---|---|---|
| ssr1 | Na10-E02 | 13 | 55 | TCGCGCATGTAATCAAAATC | TGTGACGCATCCGATCATAC |
| ssr2 | Na12–C08 | 11 | 55 | GCAAACGATTTGTTTACCCG | CGTGTAGGGTGATCTAGATGGG |
| ssr3 | Na14–C12 | 9 | 54 | CACATTTTGGTTCAATTCGG | TACGACGCTGGTTTCGATTC |
| ssr4 | Na14-D07 | 1 | 55 | GCATAACGTCAGCGTCAAAC | CTGCGGGACACATAACTTTG |
| ssr5 | Na12–C07 | 3 | 58 | ACTCAACCCCACAAACCTG | AGTTCCCCGGATCCGATTAG |
| ssr6 | Ol10-A05 | 2 | 55 | TGTAATAACCCGACCCATCC | CTCTCTCGCTCTCTCGATCC |
| ssr7 | Ol11-H02 | 4 | 55 | TCTTCAGGGTTTCCAACGAC | AGGCTCCTTCATTTGATCCC |
| ssr8 | Ol11-G11 | 13 | 54 | GTTGCGGCGAAACAGAGAAG | GAGTAGGCGATCAAACCGAG |
| ssr9 | Ol10-F11 | 11 | 55 | TTTGGAACGTCCGTAGAAGG | CAGCTGACTTCGAAAGGTCC |
| ssr10 | Ol10-H04 | 17 | 55 | TCACCCCTCTATATCCACCC | CAGAATCTGCCTGAACATCG |
| ssr11 | BRMS-040 | 1 | 55 | TCGGATTTGCATGTTCCTGAT | CCGATACACAACCAGCCAACTC |
| ssr12 | Ol10-H02 | 12 | 54 | TCGGATTTGCATGTTCCTGACT | CCGATACACAACCAGCCAACTC |
| ssr13 | Ni2-A11 | 16 | 55 | AACAAACAAGAGTCGAATACGG | AATGCCCTCTAACTGAGCCC |
| ssr14 | Na10–F06 | 3 | 55 | CTCTTCGGTTCGATCCTCG | TTTTTAACAGGAACGGTGGC |
| ssr15 | BRMA-019 | 5 | 55 | CCCAAACGCTTTTGACACAT | GGCACAATCCACTCAGCTTT |
| ssr16 | BRMS-008 | 3 | 55 | AGGACACCAGGCACCATATA | CATTGTTGTCTTGGGAGAGC |
| ssr17 | Ra2-A11 | 9 | 55 | GACCTATTTTAATATGCTGTTTTACG | ACCTCACCGGAGAGAAATCC |
| ssr18 | Ra2-E03 | 10 | 56 | AGGTAGGCCCATCTCTCTCC | CCAAAACTTGCTCAAAACCC |
| ssr19 | Ra2-E11 | 13 | 55 | GGAGCCAGGAGAGAAGAAGG | CCCAAAACTTCCAAGAAAAGC |
| ssr20 | Ra2-E12 | 8 | 55 | TGTCAGTGTGTCCACTTCGC | AAGAGAAACCCAATAAAGTAGAACC |
| ssr21 | 0l12-E03 | 7 | 55 | CTTGAAGAGCTTCCGACACC | GACGGCTAACAGTGGTGGAC |
| ssr22 | 0l12-F11 | 1 | 55 | AAGGACTCATCGTGCAATCC | GTGTCAGTGGCTACAGAGAC |
| ssr23 | Ol13-C12 | 8 | 55 | AGAGGCCAACAAAGAACACC | GAAGCAGCACCAGTGACAAG |
| ssr24 | Ol10-F11a | – | 55 | TTTGGAACGTCCGTAGAAGG | CAGCTGACTTCGAAAGGTCC |
| ssr25 | Na14–B03 | – | 55 | GATGGTCGCCGATTCAATGA | CCCATCAGCACTAGAAACCA |
| ssr26 | Na14-E08 | 14 | 55 | TTACTATCCCCTCTCCGCAC | GCGGATTATGATGACGCAG |
| ssr27 | Ol10-B01 | 17 | 55 | CCTCTTCAGTCGAGGTCTGG | AATTTGGAAACAGAGTCGCC |
| ssr28 | Ol11-H06 | 19 | 55 | TCCGAACACTCTAAGTTAGCTCC | TTCTTCACTTCACAGGCACG |
| ssr29 | Na12–C06 | 1 | 55 | AACGGATGAAGAACACATTGC | TAGGGCCTGTTATTCGATGG |
| ssr30 | Na12-D03 | 16 | 56 | GGTAAGCCAAAAACCCTTCC | GAAACCGGTAACAAAGTCGG |
| ssr31 | Na14-E02 | 13 | 56 | ACTGGCTACATGAGTTTCAGTG | GAGGGAAGACAACTGGTCTCA |
| ssr32 | Ol10-G09 | 15 | 56 | TGCTTCCTTTTTCTTCGCTC | GAAGCACGAACGCGAGAG |
| ssr33 | Na12–F12 | 13 | 56 | CGTTCTCACCTCCGATAAGC | TCCGATGTAGAATCAGCAGC |
| ssr34 | Ol11-B05 | 3 | 55 | TCGCGACGTTGTTTTGTTC | ACCATCTTCCTCGACCCTG |
| ssr35 | Ol10-H07 | 16 | 55 | TAGAGATGTCACCCGAAGGC | AGCTTCATTTCAGTCGGTGG |
| ssr36 | Ol12-B05 | 14 | 55 | GGAAAGCGAAGAGTGACGAC | ATTGGGTAAAGCTGTGCTCG |
| ssr37 | Ni2–B01 | 19 | 55 | AAGGAGATTGTTTTTGGGGC | AAGACTAATAAACACACGGCG |
| ssr38 | Ni4-E08 | 9 | 55 | GATTTTGAGGAAGCGGAGG | CAAAGCACTGAGAGAGAGAGAGAG |
| ssr39 | Ra2-F11 | 13 | 55 | TGAAACTAGGGTTTCCAGCC | CTTCACCATGGTTTTGTCCC |
| ssr40 | Ra2-G09 | 1 | 55 | ACAGCAAGGATGTGTTGACG | GATGAGCCTCTGGTTCAAGC |
| ssr41 | Ra2-E07 | 10 | 55 | ATTGCTGAGATTGGCTCAGG | CCTACACTTGCGATCTTCACC |
| ssr42 | Ol11-C02 | 17 | 55 | GCATTGCAATCTTGTTGGTC | CGTTTCCATACAGATCGTAAGAC |
| ssr43 | Ol13-C03 | 19 | 57 | GATCGGAGATGCGATGAGAG | GCATGCACCAGTGAAAAACTC |
| ssr44 | Ra2-D04 | 6 | 55 | TGGATTCTCTTTACACACGCC | CAAACCAAAATGTGTGAAGCC |
| ssr45 | Ra2-A10 | 6 | 55 | CCAGTGTGTGTGTGTGTGTG | TTTAACAGATAGCGCAGTGGTC |
| ssr46 | Ol12-G04 | 8 | 55 | CGAACATCTTAGGCCGAATC | GGTTAACCTGCGGGATATTG |
| ssr47 | Ra3-H10 | 5 | 55 | TAATCGCGATCTGGATTCAC | ATCAGAACAGCGACGAGGTC |
| ssr48 | Ra2-E04 | – | 55 | ACACACAACAAACAGCTCGC | AACATCAAACCTCTCGACGG |
| ssr49 | Ol13-E08 | 12 | 55 | TTCGCAACTCCTCCTAGAATC | AAGGTCTCACCACCGGAGTC |
| ssr50 | Ra2-A01 | 7 | 55 | TTCAAAGGATAAGGGCATCG | TCTTCTTCTTTTGTTGTCTTCCG |
| ssr51 | Ra3-D02B | 3 | 55 | CACAGGAAACCGTGGCTAGA | AACCCAACCTCAACGTCTTG |
| ssr52 | Na12–C03 | 12 | 55 | ATCGTTGCCATTAGGAGTGG | ACCAAATTAACCCTCTTTGC |
| ssr53 | Na12–H04 | 11 | 55 | TTTATCGTCTTTCCCCTCCC | ACAAGGAACTAGAGAGAGAGAG |
| ssr54 | Ol12-D05 | 18 | 47 | TCCATGACCAACGACAAGGTC | AAGAGGCGACTTCTATTGCG |
| ssr55 | Ra2-H06 | – | 47 | GAATTCAGAGGTATCTACACGGC | TAACAAAGACCCTGCGTTCC |
| ssr56 | BRMS-033 | 1 | 55 | GCGGAAACGAACACTCCTCCCATGT | CCTCCTTGTGCTTTCCCTGGAGACG |
| ssr57 | BRMS-005 | 5 | 47 | ACCTCCTGCAGATTCGTGTC | GCTGACCTTTCTTACCGCTC |
| ssr58 | BRMS-036 | 1 | 47 | GGTCCATTCCTTTTTGCATCTG | CATGGCAAGGGGTAACAAACAT |
| ssr59 | Na12–B09 | – | 47 | ACGGAAGATCAAACAGCTCC | TGAGCGACCCATTCTTTAGG |
| ssr60 | BRMS-015 | – | 47 | TCGCCAATAGAACCCAAAACTT | CATCTCCATTGCTGCATCTGCT |
| ssr61 | BRMS-027 | 1 | 47 | GCAGGCGTTGCCTTTATGTA | TCGTTGGTCGGTCACTCCTT |
| ssr62 | BRMS-042 | – | 47 | AGCTCCCGACAGCAACAAAAGA | TTCGCTTCCTTTTCTGGGAATG |
| ssr63 | Ni4–B10 | 11 | 47 | GTCCTTGAGAAACTCCACCG | CCGATCCCATTTCTAATCCC |
| ssr64 | Na10-G10 | – | 45 | TGGAAACATTGGTGTTAAGGC | CATAGATTCCATCTCAAATCCG |
| ssr65 | Ol10-G08 | 14 | 54 | TGCTTAATTGATTAGGGCAG | TTACCTCATCAGGTGGAGGC |
| ssr66 | Ra2-A04 | 5 | 55 | AAAAACTCCTCTTCAACG | CCCAAAGTTAGGTTTTAATGTAATCTC |




**(a)**




**(b)**

**Fig. 2.** Documentation gel showing polymorphic bands of SSRs amplified by a) Primer Ol10-G08 b) Ol12-G04 (K1 to K31 are thirty-one accessions of kale).

### *2.3. Statistical analysis*

#### *2.3.1. Morphological*

The datasets from two years were pooled. Statistical Tool for Agricultural Research (STAR) software was used to analyse the morphological data for descriptive statistics and principal component analysis (PCA). Program R [30] was used to determine the Pearson correlation coefficients for 24 agronomic traits.

#### *2.3.2. Molecular*

The SSR genotypic data was processed by PowerMarker v3.0 [31] to provide fundamental marker statistics values of expected heterozygosity ($H_E$), average heterozygosity ($H_{av}$), effective multiplex ratio (EMR), marker index (MI), number of alleles per locus, polymorphic information content (PIC), discriminating power (D) and resolving power (R). Using DARwin software, the unweighted neighbour joining tree was constructed based on genetic distances [32]. Program Arlequin v 3.5 [33] was used to estimate analysis of molecular variance (AMOVA).

For population structuring, model-based tool STRUCTURE 2.2 [34] was used. Using the admixture model with a burn-in of $10^5$ and a run length of $10^5$ the STRUCTURE programme was run five times for each K-value, ranging from 1 to 10. For each K, an average likelihood value [LnP (D)] was determined across all runs. In order to account for STRUCTURE's overestimation of the number of sub-groups, the most probable K value was determined using the *ad hoc* criterion ($\Delta K$) of Evanno [35] estimated by STRUCTURE HARVESTER [36]. The accessions with membership probabilities of $\geq$0.80 were assigned to subgroups using the run of estimated numbers of subgroups displaying maximum likelihood. The admixed group received accessions with membership probabilities of $<$0.80 [37].

Association mapping was carried out with TASSEL 3.0 [38]. To confirm the link between markers and traits, we employed two models-the mixed linear model (Q + K MLM) and the general linear model (Q GLM). The population structure (Q) matrix obtained from the STRUCTURE programme was used to perform the Q GLM approach. Ten thousand permutations were done to get a P-value between 0.05 and 0.01 for marker significance. The population structure Q matrix and kinship (K) matrix were employed in the Q + K MLM approach at $P < 0.05$ and $P < 0.01$. The kinship coefficient was calculated in SPAGeDi software [39]. The loci were considered to be in significant LD if $P < 0.01$.

## 3. Results

### *3.1. Morphological diversity and correlation*

The analysis of variances conducted with respect to twenty-four morphological characteristics of the germplasm revealed

**Table 3**
Analysis of variance for various agronomical traits in 62 kale genotypes.

| Source of Variation | DF | PH | CD | LBL | LBW | LBNI | LBT | NL | NLO | PeL | PeD | VSL | VSW |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Replications | 1 | 2.66 | 0.14 | 3.63 | 0.17 | 0.68 | 0.0 | 974.40 | 0.03 | 4.46 | 0.59 | 0.17 | 12.62 |
| Genotypes | 61 | 187.30* | 74.79* | 11.96* | 8.82* | 245.30* | 0.02* | 567.88* | 7.14* | 19.16* | 4.79* | 298.56* | 558.31* |
| Error | 61 | 2.04 | 3.10 | 0.67 | 0.43 | 1.56 | 0.001 | 1.53 | 0.01 | 0.74 | 0.22 | 3.80 | 8.99 |
| Source of Variation | DF | SFB | FSL | NFBPP | IF | EF | FP | SL | SW | NSS | DM | SW | LY |
| Replications | 1 | 0.06 | 0.007 | 79.04 | 3.23 | 44.16 | 20.16 | 1.11 | 0.02 | 6.78 | 28.07 | 0.30 | 16.48 |
| Genotypes | 61 | 3.69* | 0.24* | 2378439.5* | 47.49* | 103.45* | 126.55* | 2.41* | 1.16* | 35.27* | 43.23* | 2.50* | 145.96* |
| Error | 61 | 0.02 | 0.005 | 24.71 | 3.55 | 4.06 | 6.58 | 0.05 | 0.04 | 0.94 | 3.28 | 0.023 | 6.33 |

**$P < 0.01$, *df* degree of freedom, *PH* Plant height, *CD* Canopy diameter, *LBL* Leaf blade length, *LBD* Leaf blade diameter, *LBNI* Leaf blade no. of incisions, *LBT* Leaf blade thickness, *NL* No. of leaves, *Nlo* No. of lobes, *Pel* Petiole length, *Ped* Petiole diameter, *Vsl* Vegetative stem length, *Vsw* Vegetative stem width, *SFB* Size of floral buds, *Fsl* Floral stalk length, Nfbpp no. of floral buds/plant,IF Initiation of flowering, EF End of flowering, FP Flowering period, SL Siliqua length, SW siliqua width, NSS No. Of seeds/siliqua, DM Days to maturity, SW Seed weight, LY Leaf yield. Values in the cells are mean sum of squares.

significant mean squares across the accessions (Table 3). The coefficient of variation (CV) for different traits ranged from 0.29 to 17.05 (Table 4). Number of leaf lobes (17.05 %) had the largest CV, followed by leaf yield (11.01 %), vegetative stem length (8.63 %), leaf blade thickness (8.41 %), flowering period (7.49 %) and petiole length (7.40 %). Leaf yield being complex quantitative trait showed strongest correlations with leaf blade diameter (0.347**), petiole length (0.247**) and siliqua width (0.364**) at $P < 0.01$. The findings of correlation analyses are shown in Table 5.

### 3.1.1. Morphological cluster analysis

According to the results of the cluster analysis of different traits, all accessions were divided into three clusters, with cluster I, III and II having 49, 8 and 5 accessions, respectively (Fig. 3). The y-axis of the dendrogram represents distance between clusters.

### 3.1.2. Principal component analysis

The PCA showed that eigenvalues for all PCs ranged from less than 1 to 4 (Table 6). Based on Kaiser criterion, seven PCs (PC1 to PC7) were found to have eigenvalues >1, which explained a cumulative variation of 69.21 % (Table 6). However, the Scree curve tends to become linear after PC3 (Fig. 4), we considered only first three PCs which explain the cumulative variance of 42.31 % (Table 6).

The PC1 explained the highest variation of 17.04 %, followed by PC2 explaining 15.25 % and then PC3 explaining 10.02 % of variation. Thus, the selection of lines and traits based on PC1 shall be most useful. The PC1 has positive association with 17 traits *viz.,* PH, CD, LBL, LBD, PeL, PeD, VSW, SFB, FSL, NFBPP, IF, SL, SW, DM, NSS, X1000SW, and LY whereas, negatively associated with 7 traits viz., LBT, LBNI, NL, NLO, VSL, EF, and FP (Table 7).

### 3.1.3. Molecular genetic diversity

Out of 75 SSR markers, nine markers didn't amplify any product in any genotype and were excluded from analysis. Sixty-six SSR markers amplified total 269 alleles on 62 kale accessions. The average number of alleles per locus was 4.07, with the range being 1 to 11. Marker Ol12-G04, showed the highest number of alleles, whereas a group of markers *viz.*, Na12–C07, BRMS-040, Ni2-A11, Ol10-G09, Ol11-C02, Ra2-H06, BRMS-005 and Na12–B09 amplified only one allele. Expected heterozygosity ($H_E$) ranged from 0.00 (BRMS-019, Ra2-F11, Ol12-D05, BRMS-042 and Ra2-H06) to 1.887 (Ol13-C12) with an average of 0.392. Similarly, effective multiplex ratio (EMR), average heterozygosity ($H_{av}$), marker index (M), discriminatory power (D) and resolving power (R) ranged from 0.0007 (Na12–C06) to 3.403 (Ra2-E07, Na10-G10), 0.00 (BRMS-019, Ra2-F11, Ol12-D05, BRMS-042 and Ra2-H06) to 2.483 (Ol10-G09), 0.00 (Ra2-F11, Ra2-F11, BRMS-042, Ol12-D05, Ra2-H06) to 2.451 (Ol10-G09), 0.00 (Ra2-F11, BRMS-019, BRMS-042, Ol12-D05, Ra2-H06) to 2.064 (Ol10-G09) and 0.00 (BRMS-042, Ol12-D05, Ra2-H06, BRMS-019, Ra2-F11, Na14-E08, Na12–C06, Na12-D03, Na14-E02, Ol10-G09, Ra2-D04 and Ra2-A10) to 3.484 (Ol10-A05). The PIC value ranged from 0.00078 (Na12–C06) to 0.95348 (Na14-E02) and averaged to 0.407. Ten SSR loci *viz.*, (Ol13-C12, Ol11-H06, Na14-E02, Ra2-F11, Ra2-D04, Ra2-A10, RA3-Do2B, Na12–C03, Ra2-E04, Ra2-A01 had PIC values greater than 0.5, demonstrating high level of discriminating potential of the selected SSRs (Table 8).

### 3.1.4. Population structure and molecular cluster analysis

A model-based application called STRUCTURE was used to divide the accessions into proper subgroups in order to study genetic relationships among them. The largest $\Delta K$ value was apparent for all lines when $K = 3$ (Fig. 5). Accessions were separated into three

**Table 4**
Descriptive statistics of twenty-four agronomic traits across 62 kale accessions.

| S. No. | Trait | Mean ± SD | Range | Coefficient of Variation (%) |
|---|---|---|---|---|
| 1 | Plant height (cm) | 50.56 ± 9.69 | 20.80–90.60 | 2.82 |
| 2 | Canopy diameter (cm) | 53.34 ± 6.22 | 27.86–67.24 | 3.30 |
| 3 | Leaf blade length (cm) | 20.10 ± 2.51 | 14–29.20 | 4.10 |
| 4 | Leaf blade width (cm) | 14.41 ± 2.14 | 9.60–22 | 4.54 |
| 5 | Leaf blade no. of incisions | 19.27 ± 11.07 | 0.00–60 | 6.48 |
| 6 | Leaf blade thickness (cm) | 0.36 ± 0.1036 | 0.15–0.81 | 8.41 |
| 7 | Number of leaves | 30.52 ± 17.04 | 13.20–131.20 | 4.05 |
| 8 | Number of leaf lobes | 0.65 ± 1.88 | 0.00–7.40 | 17.05 |
| 9 | Petiole length (cm) | 11.65 ± 3.15 | 5.80–25.60 | 7.40 |
| 10 | Petiole diameter (cm) | 10.57 ± 1.58 | 5.73–13.77 | 4.48 |
| 11 | Vegetative stem length (cm) | 22.59 ± 12.25 | 13.80–87.80 | 8.63 |
| 12 | Vegetative stem width (cm) | 64.49 ± 16.78 | 20.06–93.71 | 4.65 |
| 13 | Size of floral buds (mm) | 7.45 ± 1.36 | 4.48–11.12 | 2.18 |
| 14 | Floral stalk length (cm) | 1.57 ± 0.35 | 1–2.77 | 4.68 |
| 15 | No. of buds/plant | 1737.14 ± 1086.08 | 299–4116 | 0.29 |
| 16 | Time of flowering (days) | 252.37 ± 5.03 | 235–267 | 0.74 |
| 17 | End of flowering (days) | 286.45 ± 7.33 | 274–302 | 0.70 |
| 18 | Flowering period (days) | 34.26 ± 8.14 | 15–60 | 7.49 |
| 19 | Siliqua length (cm) | 5.67 ± 1.11 | 2.83–9.37 | 4.15 |
| 20 | Siliqua width (cm) | 3.13 ± 0.77 | 1.40–4.82 | 6.49 |
| 21 | No. of seeds/siliqua | 16.90 ± 4.24 | 8.20–29.40 | 5.74 |
| 22 | Days to maturity | 321.85 ± 4.83 | 311–331 | 0.56 |
| 23 | 1000 seed weight (g) | 4.71 ± 1.12 | 2.10–7.80 | 3.21 |
| 24 | Leaf yield (t/ha) | 22.86 ± 8.70 | 11.13–52.94 | 11.01 |

**Table 5**
Correlation coefficient for 24 traits of 62 kale accessions.

| | PH | CD | LBL | LBD | LBNI | LBT | NL | NLO | PeL | PeD | VSL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PH | | | | | | | | | | | |
| CD | −0.373** | | | | | | | | | | |
| LBL | | 0.201* | | | | | | | | | |
| LBD | | | 0.555** | | | | | | | | |
| LBNI | −0.256** | | 0.258** | | | | | | | | |
| LBT | | −0.285** | | | | | | | | | |
| NL | | −0.388** | −0.378** | −0.304** | | 0.229* | | | | | |
| NLO | −0.535** | | | −0.306** | 0.413** | | 0.244** | | | | |
| PeL | | | 0.362** | | 0.297** | | | 0.245** | | | |
| PeD | | 0.271** | 0.273** | 0.273** | | | −0.517** | −0.341** | | | |
| VSL | 0.543** | −0.570** | −0.373** | | −0.368** | 0.295** | 0.414** | | −0.311** | −0.225* | |
| VSW | −0.226* | 0.368** | 0.298** | 0.258** | | −0.203* | −0.430** | −0.224* | | 0.209* | −0.621** |
| SFB | | | | | | | −0.193* | −0.421** | | | |
| FSL | −0.275** | 0.241** | | | | | | | | | −0.245** |
| NFBPP | | 0.266** | | −0.294** | | −0.297** | | | 0.223* | | −0.239** |
| IF | −0.206* | | | | | 0.251** | | | | 0.233** | |
| EF | | | | | | | | 0.197* | | | |
| FP | | | | | 0.203* | | | | 0.202* | | |
| SL | | | | 0.339** | | | | −0.405** | | 0.298** | |
| SW | | 0.191* | | 0.232** | −0.291** | −0.221* | | | 0.271** | | |
| NSS | | | | | | −0.206* | | −0.369** | | 0.246** | |
| DM | | | | | −0.360** | | −0.241** | −0.436** | | 0.201* | |
| WS1000 | 0.519** | | | 0.220* | −0.334** | | | −0.592** | | 0.231** | 0.177* |
| LY | | | 0.210* | 0.347** | | | | | 0.247** | | |

| | VSW | SFB | FSL | NFBPP | IF | EF | FP | SL | SW | NSS | DM | WS1000 | LY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PH | | | | | | | | | | | | | |
| PD | | | | | | | | | | | | | |
| LBL | | | | | | | | | | | | | |
| LBD | | | | | | | | | | | | | |
| LBNI | | | | | | | | | | | | | |
| LBT | | | | | | | | | | | | | |
| NL | | | | | | | | | | | | | |
| NLO | | | | | | | | | | | | | |
| PeL | | | | | | | | | | | | | |
| PeD | | | | | | | | | | | | | |
| VSL | | | | | | | | | | | | | |
| VSW | | | | | | | | | | | | | |
| SFB | 0.299** | | | | | | | | | | | | |
| FSL | | 0.181* | | | | | | | | | | | |
| NFBPP | 0.209* | 0.365** | | | | | | | | | | | |
| IF | | | | | | | | | | | | | |
| EF | | | | | | | | | | | | | |
| FP | | | | 0.226* | −0.479** | 0.776** | | | | | | | |
| SL | | 0.273** | 0.274** | | 0.186* | | | | | | | | |
| SW | | | | | | | | | | | | | |
| NSS | | | | 0.227* | | | | 0.394** | | | | | |
| DM | | | | | 0.228* | 0.217* | | 0.317** | | | | | |
| WS1000 | | | | | | | | 0.340** | 0.238** | | 0.470** | | |
| LY | | | | | | | | 0.187** | 0.364** | | | | |

**P < 0.01.

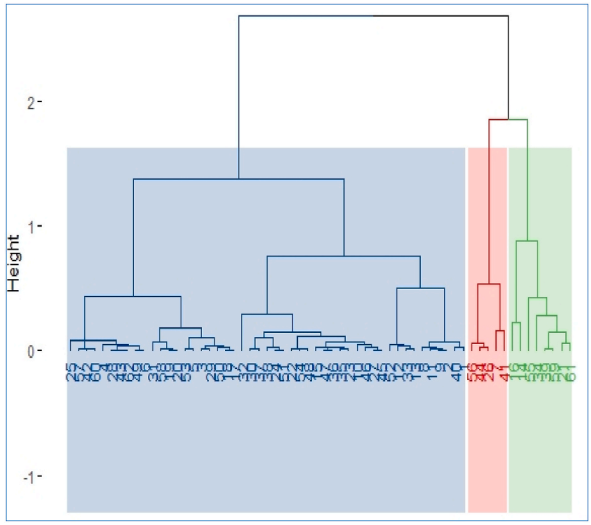



**Fig. 3.** Dendrogram showing relationship among kale genotypes using Agglomerative clustering method.

**Table 6**
Summary of principal component analysis.

| Statistics | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|---|---|---|---|---|---|---|---|
| Standard deviation | 2.0226 | 1.9129 | 1.5506 | 1.3942 | 1.3446 | 1.1999 | 1.1245 |
| Proportion of Variance | 0.1704 | 0.1525 | 0.1002 | 0.081 | 0.0753 | 0.06 | 0.0527 |
| Cumulative Proportion | 0.1704 | 0.3229 | 0.4231 | 0.5041 | 0.5794 | 0.6394 | 0.6921 |
| Eigenvalues | 4.0907 | 3.659 | 2.4044 | 1.9438 | 1.8079 | 1.4398 | 1.2646 |

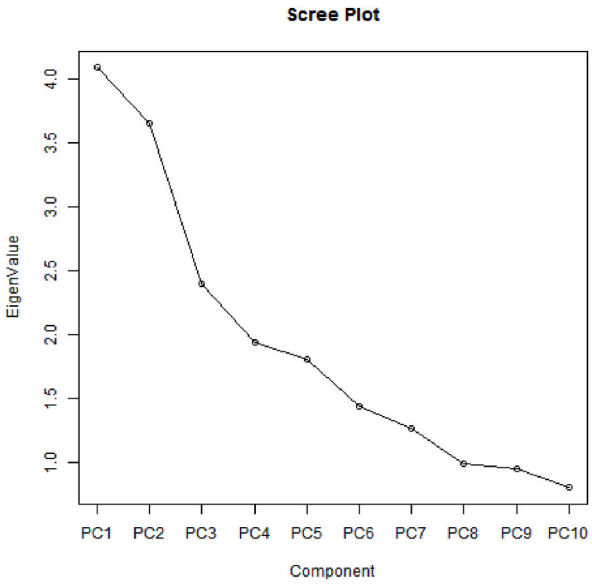



**Fig. 4.** Scree plot showing Eigen values for each principle component Fig. 4 is available only in black and white.

groups based on the membership probability criterion of 0.8 [40].

Sixty-two lines were assigned to the relevant 1–3 sub-populations using this method, accounting for 56.45 % (35), 9.68 % (6) and 27.42 % (17) of the total studied germplasm (Fig. 6). There were only four genotypes in the admixture group: CITH-KC-2, *Khanyari*, Japanese Green and NW-Sag-20R (Table 9). *Fst* mean values were found to be 0.4178 for sub-population I, 0.5892 for sub-population II (0.5892) and 0.3200 for sub-population III.

According to AMOVA, only 32.27 % genetic variance was detected among populations whereas 67.73 % was found within

**Table 7**
Eigenvectors of each trait for first three Principal Components.

| Traits | PC1 | PC2 | PC3 |
|---|---|---|---|
| PH | 0.0204 | **−0.3634** | 0.2037 |
| CD | 0.2338 | 0.2733 | 0.0986 |
| LBL | 0.2569 | 0.1952 | −0.1239 |
| LBD | 0.283 | −0.0655 | −0.2687 |
| LBT | −0.165 | −0.0796 | **−0.3299** |
| LBNI | −0.0714 | **0.3305** | 0.0235 |
| NL | **−0.3306** | −0.0513 | −0.0104 |
| NLO | **−0.2733** | **0.3372** | −0.082 |
| PeL | 0.0916 | 0.2265 | 0.0744 |
| PeD | **0.3148** | −0.071 | −0.0499 |
| VSL | −0.2585 | **−0.3653** | 0.0138 |
| VSW | **0.3021** | 0.1842 | 0.0009 |
| SFB | 0.2517 | −0.023 | 0.0981 |
| FSL | 0.0899 | 0.1384 | −0.0368 |
| NFBPP | 0.0951 | 0.1019 | **0.4299** |
| IF | 0.0274 | −0.1373 | **−0.3419** |
| EF | −0.0383 | −0.056 | 0.2974 |
| FP | −0.0919 | 0.0651 | **0.4721** |
| SL | 0.2586 | −0.1671 | −0.0423 |
| SW | 0.1695 | −0.073 | 0.0484 |
| DM | 0.1292 | −0.3055 | 0.0357 |
| NSS | 0.2212 | −0.0574 | 0.1723 |
| X1000SW | 0.1721 | −0.3293 | 0.1293 |
| LY | 0.1868 | 0.026 | −0.2469 |

population (Table 10). A neighbour joining tree (Fig. 7) was constructed where the accessions were divided into three major groups using DARwin software, which produced results that were comparable to those of the STRUCTURE analysis and morphological analysis.

### *3.1.5. Marker-trait associations (MTAs)*

On the basis of combinations of the 66 SSR loci, an LD pattern was derived. D prime values ranged from 0.48 to 0.73, with an average of 0.58 and a significant LD value of 66.40 % was detected (Table 11). We identified 110 MTAs involving 30 SSR markers and 24 agronomic traits using a Q GLM model (Supplementary Table 1). There were six MTAs involving five markers using a Q + K MLM model (Supplementary Table 2). Information on significant MTAs overlap between Q GLM and Q + K MLM at a significance level of $P \leq 0.01$ (Table 12). In both Q GLM and Q + K MLM models, one marker ssr27 was found to be linked with plant height at significance level of $P \leq 0.0001$.

## 4. Discussion

For effective conservation, management, and exploitation of genetic resources in varietal development initiatives, a thorough characterization of plant genetic resources and knowledge of the genetic relationships in the germplasm collections are of prime importance. For optimal conservation, management, and utilization in varietal development, it is essential to characterize the germplasm and comprehend genetic relationships among genotypes. Such genetically characterized germplasms then act as genetic stocks for multi-target breeding programmes. The DNA markers further help in genetic improvement *via* selecting the appropriate members of a population on the basis of their association with the desired trait, technically called marker-trait association. Therefore, combining conventional and molecular approaches provides a better insight and enhances the credibility of morphological findings and interpretations to be utilized for the breeding of any crop.

The present study comprises the phenotypic as well as molecular characterization of indigenous kale accessions developed from farmer field collections that are expected to express considerable variability. Furthermore, microsatellite or SSR markers are deemed highly reproducible DNA markers across species. Therefore, in the lack of SSR in kale, those developed in other *Brassica* species were used in the study to ascertain their usefulness in achieving the above-mentioned objectives.

### *4.1. Genetic variability and relationships*

In this study, we evaluated 62 accessions of kale indigenous to Kashmir, India. Significant differences were found for most of the morphological parameters. Leaf yield and number of lobes were found to have larger than 10 % CV, indicating the presence of significant genetic variation in the germplasm for the traits. The high CV in the characteristics is most likely caused by the fact that genotypes were collected from various local diversity hotspots. A large CV for leaf yield (11 % of 22.86 t/ha) is highly desirable, as it offers greater opportunity to improve this most important trait from growers' point of view.

Distribution of accessions among clusters on the basis of morphology grouped largest number of accessions into cluster I, which was followed by clusters III and II accounting for 79.03 %, 12.90 %, and 8.06 % of total germplasm. The clustering suggests that these kale

**Table 8**
Genetic diversity parameters based on 66 SSR markers in 62 kale accessions.

| Marker | Expected heterozygosity ($H_E$) | Average heterozygosity (*H. av*) | Effective multiplex ratio (*EMR*) | Marker index (*MI*) | Discriminatory power (*D*) | Resolving power (*R*) | Polymorphism information content (PIC) | No. of alleles (Na) |
|---|---|---|---|---|---|---|---|---|
| Na10-E02 | 0.4437 | 0.0014 | 1.6613 | 0.0024 | 0.8903 | 1.3871 | 0.3549 | 5 |
| Na12–C08 | 0.3798 | 0.0012 | 1.2742 | 0.0016 | 0.9357 | 1.5161 | 0.3813 | 5 |
| Na14–C12 | 0.2482 | 0.0013 | 0.4355 | 0.0006 | 0.9796 | 0.8710 | 0.4226 | 3 |
| Na14-D07 | 0.2248 | 0.0036 | 0.8710 | 0.0032 | 0.2433 | 0.2581 | 0.4281 | 3 |
| Na12–C07 | 0.4694 | 0.0008 | 3.3871 | 0.0028 | 0.8588 | 1.2258 | 0.3432 | 1 |
| Ol10-A05 | 0.3723 | 0.0006 | 2.2258 | 0.0015 | 0.9392 | 3.4839 | 0.3841 | 9 |
| Ol11-H02 | 0.4565 | 0.0010 | 2.4677 | 0.0026 | 0.8762 | 3.3226 | 0.3492 | 9 |
| Ol11-G11 | 0.0921 | 0.0004 | 0.1452 | 0.0001 | 0.9979 | 0.2903 | 0.4491 | 7 |
| Ol10-F11 | 0.4403 | 0.0010 | 2.2903 | 0.0023 | 0.8935 | 3.3548 | 0.3565 | 3 |
| Ol10-H04 | 0.2988 | 0.0016 | 0.5484 | 0.0009 | 0.9674 | 1.0968 | 0.4088 | 7 |
| BRMS-040 | 0.2706 | 0.0010 | 0.6452 | 0.0007 | 0.9745 | 1.2903 | 0.4168 | 1 |
| Ol10-H02 | 0.3906 | 0.0016 | 1.0645 | 0.0017 | 0.9300 | 0.3226 | 0.3143 | 4 |
| Ni2-A11 | 0.4745 | 0.0077 | 0.6129 | 0.0047 | 0.6282 | 0.7742 | 0.3619 | 1 |
| Na10–F06 | 0.4694 | 0.0025 | 1.8710 | 0.0047 | 0.6123 | 1.7419 | 0.3432 | 3 |
| BRMA-019 | 0.00 | 0.00 | 2.0000 | 0.00 | 0.00 | 0.00 | 0.4534 | 2 |
| BRMS-008 | 0.4655 | 0.0010 | 2.5806 | 0.0028 | 0.8646 | 1.4839 | 0.3450 | 7 |
| Ra2-A11 | 0.4870 | 0.0078 | 0.4194 | 0.0033 | 0.8281 | 0.8387 | 0.3348 | 5 |
| Ra2-E03 | 0.4547 | 0.0024 | 1.0484 | 0.0026 | 0.8791 | 0.8710 | 0.3500 | 3 |
| Ra2-E11 | 0.3906 | 0.0015 | 1.0645 | 0.0017 | 0.9300 | 2.1290 | 0.3771 | 4 |
| Ra2-E12 | 0.4979 | 0.0040 | 1.0645 | 0.0043 | 0.7187 | 0.3871 | 0.3294 | 2 |
| 0l12-E03 | 0.498 | 0.469 | 0.387 | 0.469 | 0.498 | 0.317 | 0.317 | 4 |
| 0l12-F11 | 0.304 | 0.318 | 0.353 | 0.318 | 0.304 | 0.378 | 0.378 | 8 |
| Ol13-C12 | 1.887 | 1.500 | 1.048 | 1.129 | 1.887 | 1.581 | 0.581 | 4 |
| Ol10-F11a | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.001 | 0.001 | 4 |
| Na14–B03 | 0.004 | 0.003 | 0.002 | 0.003 | 0.004 | 0.001 | 0.001 | 3 |
| Na14-E08 | 0.50000 | 0.49925 | 0.30934 | 0.37661 | 0.34248 | 0.00 | 0.33936 | 2 |
| Ol10-B01 | 0.33989 | 0.34027 | 0.41705 | 0.39398 | 0.40625 | 0.46489 | 0.40731 | 7 |
| Ol11-H06 | 1.00000 | 2.40323 | 1.33871 | 1.25806 | 1.09677 | 1.00000 | 0.51613 | 6 |
| Na12–C06 | 0.00403 | 0.00161 | 0.00071 | 0.00121 | 0.00110 | 0.00 | 0.00078 | 5 |
| Na12-D03 | 0.00403 | 0.00387 | 0.00095 | 0.00153 | 0.00121 | 0.00 | 0.00119 | 5 |
| Na14-E02 | 0.75203 | 0.76979 | 0.96378 | 0.93730 | 0.95244 | 0.00 | 0.95348 | 5 |
| Ol10-G09 | 0.45161 | 2.48387 | 2.29032 | 2.45161 | 2.06452 | 0.00 | 0.09677 | 1 |
| Na12–F12 | 0.4667 | 0.4667 | 1.00000 | 0.4667 | 0.8631 | 0.45161 | 0.4871 | 3 |
| Ol11-B05 | 0.4683 | 0.4683 | 1.51613 | 0.4683 | 0.8607 | 1.09677 | 0.4864 | 8 |
| Ol10-H07 | 0.4370 | 0.4370 | 1.33871 | 0.4370 | 0.8971 | 2.29032 | 0.5005 | 5 |
| Ol12-B05 | 0.4480 | 0.4480 | 2.40323 | 0.4480 | 0.8889 | 2.48387 | 0.4957 | 4 |
| Ni2–B01 | 0.4895 | 0.4895 | 1.25806 | 0.4895 | 0.8183 | 2.45161 | 0.4763 | 2 |
| Ni4-E08 | 0.4936 | 0.4936 | 1.09677 | 0.4936 | 0.6914 | 2.06452 | 0.4742 | 2 |
| Ra2-F11 | 0.00 | 0.00 | 1.00000 | 0.00 | 0.00 | 0.00 | 0.5960 | 5 |
| Ra2-G09 | 0.4752 | 0.4752 | 1.14516 | 0.4752 | 0.6265 | 1.70968 | 0.4832 | 5 |
| Ra2-E07 | 0.5000 | 0.5000 | 3.40323 | 0.5000 | 0.7476 | 2.41935 | 0.4711 | 3 |
| Ol11-C02 | 0.4870 | 0.4870 | 1.17742 | 0.4870 | 0.8249 | 1.83871 | 0.4775 | 1 |
| Ol13-C03 | 0.4520 | 0.4520 | 1.24194 | 0.4520 | 0.8816 | 0.48387 | 0.4939 | 4 |
| Ra2-D04 | 0.2835 | 0.2835 | 2.00000 | 0.2835 | 0.9712 | 0.00 | 0.5559 | 4 |
| Ra2-A10 | 0.3584 | 0.3584 | 0.498 | 0.3584 | 0.9460 | 0.00 | 0.5318 | 3 |
| Ol12-G04 | 0.4594 | 0.4594 | 0.317 | 0.4594 | 0.8728 | 0.45161 | 0.4905 | 11 |
| Ra3-H10 | 0.4480 | 0.4480 | 0.387 | 0.4480 | 0.8865 | 1.09677 | 0.4957 | 5 |
| Ra2-E04 | 0.4209 | 0.4209 | 0.469 | 0.4209 | 0.9105 | 2.29032 | 0.5075 | 5 |
| Ol13-E08 | 0.4917 | 0.4917 | 0.469 | 0.4917 | 0.8144 | 2.48387 | 0.4752 | 5 |
| Ra2-A01 | 0.4332 | 0.4332 | 0.498 | 0.4332 | 0.9006 | 2.45161 | 0.5022 | 5 |
| Ra3-D02B | 0.3122 | 0.3122 | 0.317 | 0.3122 | 0.9651 | 2.06452 | 0.5473 | 4 |
| Na12–C03 | 0.3314 | 0.3314 | 0.498 | 0.3314 | 0.9569 | 0.00002 | 0.5411 | 6 |
| Na12–H04 | 0.4480 | 0.4480 | 0.317 | 0.4480 | 0.8865 | 0.45161 | 0.4957 | 3 |
| Ol12-D05 | 0.00 | 0.00 | 2.00000 | 0.00 | 0.00 | 0.00 | 0.46489 | 2 |
| Ra2-H06 | 0.00 | 0.00 | 2.00000 | 0.00 | 0.00 | 0.00 | 0.46489 | 1 |
| BRMS-033 | 0.50000 | 0.00403 | 1.00000 | 0.00403 | 0.75203 | 0.45161 | 0.33989 | 3 |
| BRMS-005 | 0.33936 | 0.00078 | 1.51613 | 0.00119 | 0.95348 | 1.09677 | 0.40731 | 1 |
| BRMS-036 | 0.30934 | 0.00071 | 1.33871 | 0.00095 | 0.96378 | 2.29032 | 0.41705 | 3 |
| Na12–B09 | 0.49925 | 0.00161 | 2.40323 | 0.00387 | 0.76979 | 2.48387 | 0.34027 | 1 |
| BRMS-015 | 0.37661 | 0.00121 | 1.25806 | 0.00153 | 0.93730 | 2.45161 | 0.39398 | 3 |
| BRMS-027 | 0.34248 | 0.00110 | 1.09677 | 0.00121 | 0.95244 | 2.06452 | 0.40625 | 3 |
| BRMS-042 | 0.00 | 0.00 | 1.00000 | 0.00 | 0.00 | 0.00 | 0.46489 | 3 |
| Ni4–B10 | 0.47202 | 0.00254 | 1.14516 | 0.00291 | 0.85557 | 1.70968 | 0.35349 | 5 |
| Na10-G10 | 0.48887 | 0.00099 | 3.40323 | 0.00335 | 0.81953 | 2.41935 | 0.34540 | 4 |
| Ol10-G08 | 0.36006 | 0.00116 | 1.17742 | 0.00137 | 0.94513 | 1.83871 | 0.40007 | 6 |
| Ra2-A04 | 0.42817 | 0.00173 | 1.24194 | 0.00214 | 0.90447 | 0.48387 | 0.37323 | 2 |

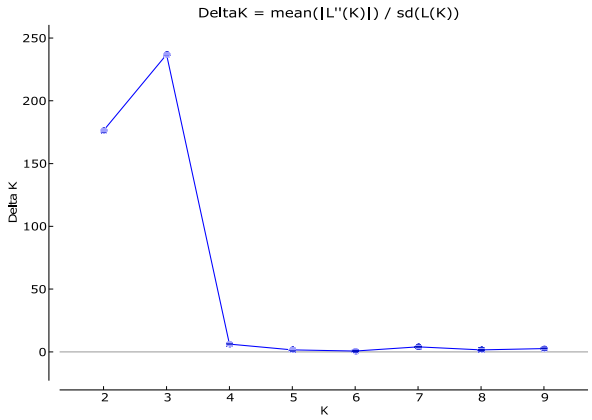



**Fig. 5.** Graphical depiction of population structure (A vertical line indicating membership in subgroup 1 (red), 2 (green), and 3 (blue) is displayed for each genotype. Genotypes are organised into K = 3 clusters according to the predicted membership coefficients. Rate of change of the likelihood distribution calculated as Delta K). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

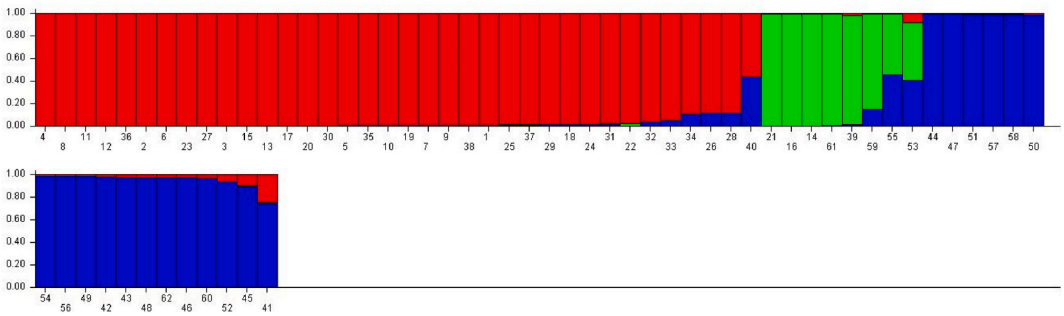



**Fig. 6.** Population structure of kale germplasm based on SSR data.

accessions have sufficient levels of genetic variation. However, the relationship observed using morphological traits alone may not reflect the underlying genetic diversity. Therefore, analysis using molecular markers was also done to comprehend genetic differences among accessions.

SSR markers based molecular characterization of these accessions was carried out by using neighbour clustering method (Fig. 7). Cluster I (red) consisted of majority of the Kashmiri collections (36) and landraces *GM Dari* and *Hanz Haaq*, cluster II (green) consisted of eight collections, mostly non-indigenous and a landrace *Kawdari*. Cluster III (blue) consisted of 18 accessions, all representing indigenous Kashmiri collections and a landrace *Khanyari*. Thus, it clearly shows that each subgroup differs from the other. The "green" group is the most diverse, as it contains mostly non-indigenoua accessions (Himachal Pradesh, Private farms). The "blue" group is the least diverse because all its members are local accessions (Fig. 7).

Agglomerative clustering (Fig. 3) of the germplasm on the basis of morphology also seems to be considerably similar to that of SSR based neighbour joining tree. Landraces *GM Dari* and *Hanz Haaq* can be seen together in the largest cluster containing most of the indigenous accessions as they are in red cluster of neighbours joining tree, which is also the largest and hosts mostly the local kashmiri accessions and a few exotics from Himachal Pradesh. Similarly, landrace *Khanyari* is in a separate cluster from other landraces in both agglomerative clustering and SSR based neighbour joining tree. With respect to the third landrace *Kawdari,* interestingly it belongs to the clusters that are almost identical in the two cases. Both cluster II of Agglomerative clustering method and cluster 'green' of SSR based tree are identical with the exception of only one accession. Such concordance between morphological clustering, molecular clustering and geographical origins of the accessions indicates the accuracy of selected SSRs in truthfully characterizing kale germplasm.

Several studies have successfully used microsatellite markers to find molecular genetic variation in *Brassica* crops [41–44]. In the present study, degree of variation among accessions was examined using the diversity indices Na, H.av, He, EMP, MI, DP, and RP. The level of variability and allele diversity are indicated by the total number of alleles in the population. Here, the numbers of alleles varied from one to eleven per locus, and in total 269 alleles were amplified. This is very large number of alleles given the fact that SSR used are from different species. Using 12 SSRs on 25 accessions of *B. oleracea* including kale, El-Esawi et al., 2016 [41] obtained only 47 alleles indicating that using larger number of *Brassica* SSR markers on *B oleracea* varieties, as done in our study, will be beneficial in case of

**Table 9**
Probability-based assignment of individuals to the (K) sub populations.

| Genotype | K1 | K2 | K3 | Sub-population assigned |
|---|---|---|---|---|
| G1 | 0.991 | 0.001 | 0.008 | 1 |
| G2 | 0.995 | 0.001 | 0.004 | 1 |
| G3 | 0.995 | 0.001 | 0.004 | 1 |
| G4 | 0.997 | 0.001 | 0.002 | 1 |
| G5 | 0.992 | 0.001 | 0.006 | 1 |
| G6 | 0.995 | 0.001 | 0.004 | 1 |
| G7 | 0.989 | 0.006 | 0.005 | 1 |
| G8 | 0.997 | 0.001 | 0.002 | 1 |
| G9 | 0.98 | 0.008 | 0.012 | 1 |
| G10 | 0.985 | 0.004 | 0.011 | 1 |
| G11 | 0.996 | 0.001 | 0.003 | 1 |
| G12 | 0.996 | 0.001 | 0.003 | 1 |
| G13 | 0.992 | 0.001 | 0.008 | 1 |
| G14 | 0.005 | 0.994 | 0.001 | 2 |
| G15 | 0.994 | 0.002 | 0.004 | 1 |
| G16 | 0.002 | 0.997 | 0.001 | 2 |
| G17 | 0.995 | 0.001 | 0.005 | 1 |
| G18 | 0.984 | 0.001 | 0.014 | 1 |
| G19 | 0.981 | 0.001 | 0.018 | 1 |
| G20 | 0.99 | 0.001 | 0.009 | 1 |
| G21 | 0.001 | 0.998 | 0.001 | 2 |
| G22 | 0.975 | 0.022 | 0.003 | 1 |
| G23 | 0.996 | 0.001 | 0.003 | 1 |
| G24 | 0.988 | 0.001 | 0.011 | 1 |
| G25 | 0.986 | 0.001 | 0.013 | 1 |
| G26 | 0.922 | 0.001 | 0.077 | 1 |
| G27 | 0.996 | 0.001 | 0.003 | 1 |
| G28 | 0.907 | 0.001 | 0.092 | 1 |
| G29 | 0.986 | 0.001 | 0.013 | 1 |
| G30 | 0.99 | 0.001 | 0.009 | 1 |
| G31 | 0.968 | 0.001 | 0.031 | 1 |
| G32 | 0.972 | 0.001 | 0.026 | 1 |
| G33 | 0.949 | 0.001 | 0.05 | 1 |
| G34 | 0.909 | 0.003 | 0.088 | 1 |
| G35 | 0.986 | 0.004 | 0.01 | 1 |
| G36 | 0.995 | 0.001 | 0.004 | 1 |
| G37 | 0.987 | 0.005 | 0.008 | 1 |
| G38 | 0.985 | 0.001 | 0.013 | 1 |
| G39 | 0.026 | 0.964 | 0.01 | 2 |
| G40 | 0.577 | 0.001 | 0.422 | Admixture |
| G41 | 0.255 | 0.005 | 0.74 | Admixture |
| G42 | 0.016 | 0.001 | 0.983 | 3 |
| G43 | 0.012 | 0.001 | 0.987 | 3 |
| G44 | 0.002 | 0.001 | 0.997 | 3 |
| G45 | 0.111 | 0.001 | 0.888 | 3 |
| G46 | 0.026 | 0.001 | 0.973 | 3 |
| G47 | 0.01 | 0.001 | 0.989 | 3 |
| G48 | 0.035 | 0.001 | 0.964 | 3 |
| G49 | 0.005 | 0.001 | 0.994 | 3 |
| G50 | 0.006 | 0.001 | 0.993 | 3 |
| G51 | 0.008 | 0.001 | 0.99 | 3 |
| G52 | 0.077 | 0.003 | 0.92 | 3 |
| G53 | 0.013 | 0.496 | 0.492 | Admixture |
| G54 | 0.007 | 0.003 | 0.99 | 3 |
| G55 | 0.003 | 0.533 | 0.465 | Admixture |
| G56 | 0.006 | 0.003 | 0.991 | 3 |
| G57 | 0.007 | 0.004 | 0.989 | 3 |
| G58 | 0.008 | 0.002 | 0.99 | 3 |
| G59 | 0.002 | 0.823 | 0.174 | 2 |
| G60 | 0.031 | 0.001 | 0.968 | 3 |
| G61 | 0.002 | 0.99 | 0.009 | 2 |
| G62 | 0.028 | 0.002 | 0.97 | 3 |

kale. The expected heterozygosity (He) differed from 0.00 to 1.887 with an average of 0.392, which was higher than the mean heterozygosity (0.266). The bulk of the reported homozygotes may actually be heterozygotes, where one allele is present but the other is absent. Similar to our results the value of average He of 0.33 has been reported by Pelc et al. [45]. SSR markers were, therefore, found to be more effective in assessing genetic diversity and generated mean values for EMR, MI, DP, and RP of 1.177, 0.244, 0.751, and 1.155, respectively [46]. PIC values ranged from 0.00078 to 0.953, with an average of 0.407, indicating the high applicability of these

**Table 10**
Analysis of molecular variance (AMOVA) for 62 kale accessions.

| Population | Sources of variation | df | Sum of squares | Estimated variability | Percentage of variation (%) | P value |
|---|---|---|---|---|---|---|
| | Among populations | 2 | 385.805 | 10.43[a] | 32.27 | <0.05 |
| | Within population | 57 | 1248.812 | 21.90[b] | 67.73 | <0.05 |
| | Total | 59 | 1634.61 | 32.34 | | |
| | *Fst* | 0.4178 | | | | |
| | | 0.5892 | | | | |
| | | 0.3200 | | | | |

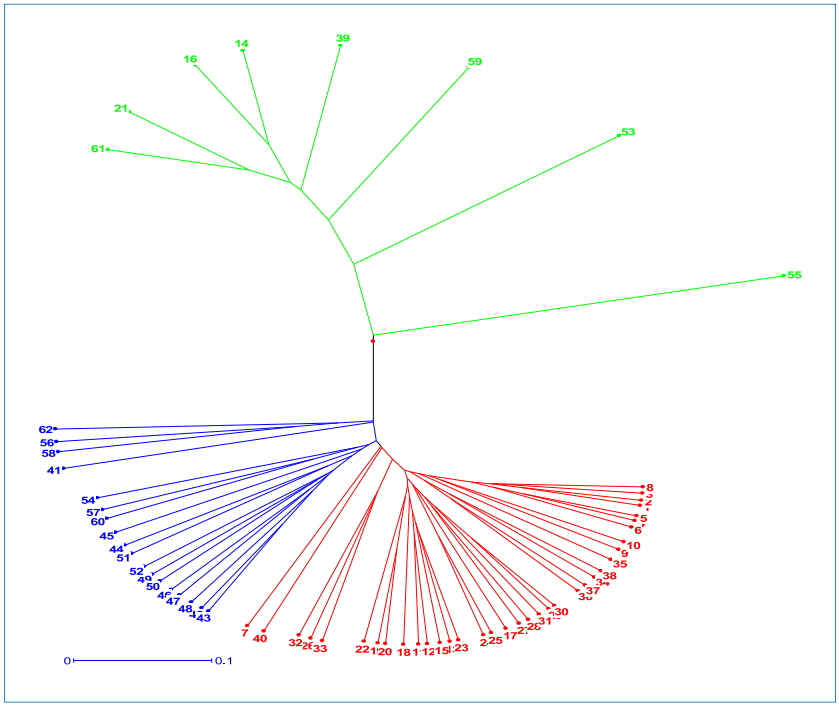



**Fig. 7.** Neighbour joining tree clustering of 62 genotypes of kale based on SSR data using DARwin.

**Table 11**
The $r^2$, D prime and LD statistics of 62 kale accessions.

| Chromosome | Mean of $r^2$ | Mean of D prime | LD (%) |
|---|---|---|---|
| 1 | 0.13 | 0.64 | 38.60 |
| 2 | 0.03 | 0.50 | 55.98 |
| 3 | 0.07 | 0.50 | 50.15 |
| 4 | 0.04 | 0.49 | 53.53 |
| 5 | 0.04 | 0.61 | 66.17 |
| 6 | 0.03 | 0.60 | 64.39 |
| 7 | 0.08 | 0.53 | 43.35 |
| 8 | 0.06 | 0.52 | 49.97 |
| 9 | 0.11 | 0.63 | 39.88 |
| 10 | 0.17 | 0.73 | 30.17 |
| 11 | 0.09 | 0.61 | 39.80 |
| 12 | 0.04 | 0.62 | 56.11 |
| 13 | 0.05 | 0.54 | 52.65 |
| 14 | 0.12 | 0.68 | 45.41 |
| 15 | – | – | – |
| 16 | 0.03 | 0.48 | 58.54 |
| 17 | 0.04 | 0.54 | 58.11 |
| 18 | 0.14 | 0.68 | 46.97 |
| 19 | 0.02 | 0.56 | 66.40 |

**Table 12**
Details on the overlapping significant marker-trait association between two models.

| Chromosome | Marker | Trait | GLM | MLM | Average P value |
|---|---|---|---|---|---|
| 4 | ssr7 | Plant height | ** | ** | ** |
| 17 | ssr27 | Plant height | *** | *** | *** |
| 16 | ssr13 | Petiole diameter | *** | ** | ** |
| 11 | ssr9 | Flowering period | * | ** | * |
| 1 | ssr29 | Siliqua length | *** | ** | ** |
| 11 | ssr9 | Seed weight | *** | ** | ** |

markers and suggest that the majority of the markers allowed for a high level of polymorphism. Our results were in conformity with the earlier findings [7,47].

Clustering generated through DARwin software using unweighted neighbour joining method was found to be the most effective in the current study for the interpretation of findings for determining the phylogenetic relationships between accessions and the closest and most distant genotypes from one another. All 62 genotypes were grouped into three main clusters/sub-populations using both an unweighted neighbour joining-based dendrogram and population structure analysis. Cluster III consisted of all the introduced accessions and two indigenous accessions. In cluster II, only one landrace '*Khanyari'* was spotted, the remaining 17 were indigenous collections. All other landraces *Kawdari*, *Hanz haaq*, and *GM Dari,* despite being phenotypically strikingly different clustered under Cluster I along with rest of the indigenous collections. The results of the structure analysis shared a similar pattern of relationship with slight mixing of genotypes. Previous studies have also observed genetic clustering based on marker data. Singh et al. [48] used the neighbour joining approach to divide 87 *Brassica juncea* genotypes into 2 groups. Likewise, Pelc et al. [45] divided the collard (*Brassica oleracea* L. var. *viridis*) landraces into 3 clusters.

In population structure studies under the present investigation, population differentiation measurements (Fst) ranged from 0.32 (sub-population III) to 0.58 (sub-population II), which are relatively high and confirm the separation of all the sub-populations,. El-Esawi et al. [41] used 12 SSRs and observed an Fst value of 0.110 for genetic differentiation in kale accessions. The fact that we used much larger number of markers in our study may be the reason why we were able to achieve higher Fst values. Alternative reason can be a relatively strong genetic difference between genotypes explored in this study. The AMOVA showed that genetic diversity within subpopulation was higher than among subpopulations. Gene flow from other varieties nearby such as by cross-pollination between fields and/or seed exchange between farmers across the valley may be the cause of higher genetic variation within the sub-populations of Kashmiri kale [7]. Nevertheless, gene flow between subpopulation is also considerable given that more than 32 per cent genetic variation is due to among population differences.

There is an apprehension that population structure analysis may produce false positive findings. Therefore, to overcome this problem, we also did association mapping analysis by using general linear model (GLM) and mixed linear model (MLM). The MLM is able to address the problem of false positives in marker-trait association by accounting for both kinship (K-matrix) and population structure (Q-matrix), whereas the GLM only considers (Q-matrix) population structure [49,50]. For 24 agronomic traits involving 30 markers, we found 110 marker-trait associations based on the Q GLM model. The Q + K MLM model led to the identification of six marker-trait associations involving five markers. The GLM approach confirmed each and every association that the MLM approach had shown to be statistically significant ($P \leq 0.001$). There have been numerous other association mapping studies using different sample sizes in Brassicaceae family [51,52]. LD levels in our study enabled us to identify marker-phenotype relationships in the inbred lines. We observed two markers associated with same trait and same marker associated with two traits. Specifically, Ol11-H02 and Ol10-B01, located on chromosome 4 and 17 of *B. oleracea* associated with same trait 'plant height'; Ol10-F11 located on chromosome 9 of *B. oleracea* has association with flowering period and seed weight; Ni2-A11 located on chromosome 16 of *B. nigra* associated with petiole diameter and Na12–C06 located on chromosome 1 of *B. napus*, associated with siliqua length, respectively. Several plant association studies for various phenotypic traits, including flowering time [26], leaf and plant architecture [21,25] and fruit quality [53] have been similarly reported. Markers Na12–C06 of *B. napus* and Ol10-F11 of *B. oleracea* found to be associated with siliqua length and seed weight, respectively, in our study are of tremendous value in breeding programmes aimed at high leaf productivity since these traits were also found to have significantly positive correlation with leaf yield. Similar is the recommendation for Ol10-F11 marker from *B. oleracea* with respect to seed yield, as seed weight associated with this marker directly contributes to seed yield. Leaf and seed productivity, being unarguably the foremost traits of agricultural and commercial importance may significantly benefit from using these markers in kale breeding programmes.

## 5. Conclusion

This is the first study on morphological and molecular diversity assessment, population structure and MTA for agronomically important traits of Indian kale. We detected significant polymorphism at both morphological and molecular levels. The morphological evaluation of twenty-four metric traits has revealed significant CV for all traits and AMOVA revealed significant genetic diversity within and among populations. High level of diversity in the studied germplasm panel showed its potential application in kale breeding programs. The PCA revealed 7 principal components that explained 69.21 % variance in the germplasm. Ten of the 66 *Brassica* SSR markers selected for this study have shown PIC values between 0.5 and 0.953 suggesting their high potential in assessing molecular genetic diversity in kale.

The molecular clustering analysis has divided the whole germplasm into 3 sub-populations and a group of 4 admixtures. One of the popular landraces of Kashmir '*Khanyari*' has placed itself in cluster II while rest of the three '*Kawdari, 'GM Dari'* and '*Hanz haaq'* have gone together into a different cluster (cluster I). As *Khanyari* is phenotypically very similar to *Kawdari* and on the other hand *Kawdari*, *GM Dari* and *Hanz haaq* are completely different from each other (Fig. 8), this finding reinforces the importance of DNA markers in genetic diversity assessment of crop species. The DNA markers can assess the genetic diversity more precisely in case of kale, as the ambiguity caused by morphology has been put at bay.

Considerable similarity between morphological and molecular clustering of the germplasm along with their consonance with geographical origins of the accessions was also observed in this study. This suggests strong candidature of the studied SSRs in diversity studies of kale.

The study also identified 110 MTAs involving 30 SSRs and 24 traits of kale. Six of these markers showed overlapping significant MTAs under GLM and MLM for 5 different traits. One of these markers, Ol10-B01 was found to associate with plant height at $p \leq 0.0001$. This marker, therefore, shows promise in marker-assisted breeding (MAB) for plant architecture manipulation. Two markers Ol10-F11 and Na12–C06 were found to be significantly associated with seed weight and siliqua length, respectively. Both these traits showed significant positive correlation with leaf yield. Secondly, seed weight can be a direct contributor to seed yield. Therefore, both leaf yield and seed yield of kale may be improved through MAB by utilizing these markers.

## Data availability statement

The data generated in the study is included in the article and supplementary material.

## Statements and declarations

.


**Fig. 8.** Kashmiri kale landraces.

## Funding

This work was supported by ICAR-Central Institute of Temperate Horticulture, Srinagar, J&K, India.

## CRediT authorship contribution statement

**Geetika Malik:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Asma Jabeen:** Writing – original draft, Formal analysis, Data curation. **Javid Iqbal Mir:** Resources, Methodology. **Rafiq Ahmad Shah:** Formal analysis. **Mohd Abas Shah:** Formal analysis. **Vishal Dinkar:** Writing – review & editing. **Muneer Ahmad Sheikh:** Resources. **Ravinder Kumar:** Formal analysis. **Om Chand Sharma:** Visualization, Project administration. **Mahendra Kumar Verma:** Writing – review & editing, Visualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

The authors duly acknowledge the funds provided by Central Institute of Temperate Horticulture, Srinagar (Indian Council of Agricultural Research, New Delhi).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e29521.

## References

[1] M. Mabry, S.D. Turner-Hissong, E.Y. Gallagher, M.A. Gore, J. Labate, et al., The evolutionary history of wild, domesticated, and feral *Brassica oleracea* (Brassicaceae), Mol. Biol. Evol. 38 (10) (2021) 4419–4434, https://doi.org/10.1093/molbev/msab183.
[2] M.W. Farnham, E.H. Davis, J.T. Morgan, J.P. Smith, Neglected landraces of collard (*Brassica oleracea* L. var. *viridis*) from the Carolinas (USA), Genet. Resour. Crop Evol. 55 (2008) 797–801, https://doi.org/10.1007/s10722-007-9284-8.
[3] J.W. Fahey, Brassica, in: C. Benjamin (Ed.), Encyclopedia of Food Sciences and Nutrition, Academic Press, 2003, pp. 606–615.
[4] A. Balkaya, R. Yanmaz, Promising kale (*Brassica oleracea* var. *acephala*) populations from Black Sea region, Turkey, N. Z. J. Crop Hortic. Sci. 33 (2005) 1–7, https://doi.org/10.1080/01140671.2005.9514324.
[5] M.E. Cartea, A. Picoaga, P. Soengas, A. Ordás, Morphological characterization of kale populations from northwestern Spain, Euphytica 29 (2003) 25–32, https://doi.org/10.1023/A:1021576005211.
[6] L. Giorgetti, G. Giorgi, E. Cherubini, P.G. Gervasi, C.M. Della Croce, V. Longo, L. Bellani, Screening and identification of major phytochemical compounds in seeds, sprouts and leaves of Tuscan black kale *Brassica oleracea* (L.) ssp *acephala* (DC) var. *sabellica* L, Nat. Prod. Res. 23 (2017) 1–10, https://doi.org/10.1080/14786419.2017.1392953.
[7] S. Christensen, R. von Bothmer, G. Poulsen, L. Maggioni, M. Phillip, B.A. Andersen, AFLP analysis of genetic diversity in leafy kale (*Brassica oleracea* L. convar. *acephala* (DC.) Alef.) landraces, cultivars and wild populations in Europe, Genet Res Crop Evol 58 (2011) 657–666, https://doi.org/10.1007/s10722-010-9607-z.
[8] D. Šamec, B. Urlić, B. Salopek-Sondi, Kale (*Brassica oleracea* var. *acephala*) as a superfood: review of the scientific evidence behind, Crit. Rev. Food Sci. Nutr. 20 (2018) 1–37, https://doi.org/10.1080/10408398.2018.1454400.
[9] P. Velasco, M.E. Cartea, C. Gonzalez, M. Vilar, A. Ordás, Factors affecting the glucosinolate content of kale (*Brassica oleracea* Acephala group), J. Agric. Food Chem. 55 (2007) 955–962, https://doi.org/10.1021/jf0624897.
[10] W. Megan, What are the health benefits of kale? Med. News Today (2020). https://www.medicalnewstoday.com/articles/270435.php.
[11] L. Bryan, How avocados and kale became so popular, BBC Work Life (2020). https://www.bbc.com/worklife/article/20190304-how-avocados-and-kale-became-so-popular.
[12] B. Sultana, F. Anwar, Flavonols (kaempeferol, quercetin, myricetin) contents of selected fruits, vegetables and medicinal plants, Food Chem. 108 (3) (2008) 879–884, https://doi.org/10.1016/j.foodchem.2007.11.053.
[13] J.M. Calderón-Montaño, E. Burgos-Morón, C. Pérez-Guerrero, M. López-Lázaro, A review on the dietary flavonoid kaempferol, Mini Rev. Med. Chem. 11 (4) (2011) 298–344, https://doi.org/10.2174/138955711795305335.
[14] S. Mali, S. Rathod, N. Kale, N. Shinde, Overview of nutraceuticals, Asian J Pharm Res 12 (1) (2022) 61–70, https://doi.org/10.52711/2231-5691.2022.00010.
[15] N. Satheesh, S.W. Fanta, Kale: review on nutritional composition, bio-active compounds, anti-nutritional factors, health beneficial properties and value-added products, Cogent Food Agric. 6 (1) (2020) 1811048, https://doi.org/10.1080/23311932.2020.1811048.
[16] N. Saroj, K. Prasad, S.K. Singh, V. Kumar, S. Maurya, P. Maurya, R. Kumar, Characterization of bioactive and fruit quality compounds of promising mango genotypes grown in Himalayan plain region, PeerJ 11 (2023) e15867, https://doi.org/10.7717/peerj.15867.
[17] N. Saroj, K. Prasad, Assessment of Himalayan plain mango genotypes for phytochemicals, biochemical-nutraceutical characterisation and quality change during storage life, Int. J. Food Sci. Technol. 58 (2023) 3781–3799, https://doi.org/10.1111/ijfs.16480.
[18] C. Lotti, P. Iovieno, I. Centomani, A.R. Marcotrigiano, V. Fanelli, G. Mimiola, et al., Genetic, bio-agronomic, and nutritional characterization of kale (*Brassica oleracea* L. var. Acephala) diversity in Apulia, Southern Italy, Diversity 10 (2) (2018) 25, https://doi.org/10.3390/d10020025.
[19] G. Miah, M.Y. Rafii, M.R. Ismail, A.B. Puteh, H.A. Rahim, K.N. Islam, M.A. Latif, A review of microsatellite markers and the applications in rice breeding programs to improve blast disease resistance, Int. J. Mol. Sci. 14 (2013) 22499–22528, https://doi.org/10.3390/ijms141122499.
[20] M. Ilyas, G. Shabbir, S.A. Jan, M.A. Rabbani, S.I. Malik, M. Ansar, et al., Estimation of linkage disequilibrium and population structure among *Brassica napus* genotypes for association mapping, Fresenius Environ. Bull. 27 (9) (2018) 5980–5986.

[21] Z. Wei, G. Zhang, Q. Du, J. Zhang, B. Li, D. Zhang, Association mapping for morphological and physiological traits in *Populussimonii*, BMC Genet. 15 (Suppl 1) (2014) S3, https://doi.org/10.1186/1471-2156-15-S1-S3.
[22] P. Tripodi, S. Soler, G. Campanelli, et al., Genome wide association mapping for agronomic, fruit quality, and root architectural traits in tomato under organic farming conditions, BMC Plant Biol. 21 (2021) 481, https://doi.org/10.1186/s12870-021-03271-4.
[23] S. Sivakumar, M. Dhasarathan, A. Karthikeyan, P. Bharathi, N.K. Vinodhana, K. Ganesamurthy, N. Senthil, Population structure and association mapping studies for yield-related traits in Maize (*Zea mays* L, Curr Plant Biol 18 (2019) 100103, https://doi.org/10.1016/j.cpb.2019.04.001.
[24] U.K. Reddy, L. Abburi, V.L. Abburi, T. Saminathan, et al., Genome-wide scan of selective sweeps and association mapping of fruit traits using microsatellite markers in watermelon, J. Hered. 106 (2) (2015) 166–176, https://doi.org/10.1093/jhered/esu077.
[25] J. Ma, A. Shi, B. Mou, M. Evans, J.R. Clark, Association mapping of leaf traits in spinach (*Spinacia oleracea* L, Plant Breed. 135 (2016) 399–404, https://doi.org/10.1111/pbr.12369.
[26] J. Zhao, M.J. Paulo, D. Jamar, P. Lou, F. van Eeuwijk, Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*, Genome 50 (2007) 963–973, https://doi.org/10.1139/G07-078.
[27] R.W. Allard, Predictive methods for germplasm identification, in: H.T. Stalker, J.P. Murphy (Eds.), Plant Breeding in the 1990s, CAB International, Wallingford Oxon, UK, 1992, pp. 119–146.
[28] P.R. Talukdar, S. Rathi, K. Pathak, S.K. Chetia, R.N. Sarma, Population structure and marker-trait association in indigenous aromatic rice, Rice Sci. 24 (3) (2017) 145–154, https://doi.org/10.1016/j.rsci.2016.08.009. ISSN 1672-6308.
[29] J.J. Doyle, J.L. Doyle, A rapid DNA isolation procedure to small amounts of fresh leaf tissue, Phytochem Bull 19 (1) (1987) 11–15.
[30] B.K. Hackenberger, R software: unfriendly but probably the best, Croat. Med. J. 61 (1) (2020) 66–68.
[31] K. Liu, S.V. Muse, Power Marker: an integrated analysis of environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129, https://doi.org/10.1093/bioinformatics/bti282.
[32] X. Perrier, J.P. Jacquemoud-Collet, Darwin software. http://darwin.cirad.fr/darwin, 2006.
[33] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol Ecol Res 10 (2010) 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x.
[34] J.K. Pritchard, W. Wen, Documentation for STRUCTURE Software, 2003, Version 2.
[35] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620, https://doi.org/10.1111/j.1365-294x.2005.02553.x.
[36] D.A. Earl, B.M. vonHoldt, Structure HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conservation Genet Resour 4 (2012) 359–361, https://doi.org/10.1007/s12686-011-9548-7.
[37] B. Stich, A.E. Melchinger, M. Frisch, H.P. Maurer, M. Heckenberger, J.C. Reif, Linkage disequilibrium in European elite maize germplasm investigated with SSRs, Theor. Appl. Genet. 111 (2005) 723–730, https://doi.org/10.1007/s00122-005-2057-x.
[38] P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635, https://doi.org/10.1093/bioinformatics/btm308.
[39] O.J. Hardy, X. Vekemans, SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels, Mol. Ecol. Notes 2 (2002) 618–620, https://doi.org/10.1046/j.1471-8286.2002.00305.x.
[40] V.V. Nachimuthu, R. Muthurajan, S. Duraialaguraja, et al., Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza sativa*, Rice 8 (1) (2015) 1–25, https://doi.org/10.1186/s12284-015-0062-5.
[41] M.A. El-Esawi, K. Germaine, P. Bourke, R. Malone, Genetic diversity and population structure of *Brassica oleracea* germplasm in Ireland using SSR markers, C R Biol 339 (2016) 133–140, https://doi.org/10.1016/j.crvi.2016.02.002.
[42] M. El-Esawi, K. Germaine, R. Malone, Assessing the genetic diversity and relationships in Irish Brassica oleracea species based on microsatellites markers, in: Proceedings of the Fifth Saudi Science Conference, Umm Al-Qura University, Saudi Arabia, 2012, April.
[43] S. Louarn, A.M. Torp, I.B. Holme, S.B. Andersen, B.D. Jensen, Database derived microsatellite markers (SSRs) for cultivar differentiation in Brassica oleracea, Genet. Resour. Crop Evol. 54 (2007) 1717–1725.
[44] X.J. Leroy, K. Leon, M. Branchard, Characterisation of Brassica oleracea L. by microsatellite primers, Plant Systemat. Evol. 225 (2000) 235–240.
[45] S.E. Pelc, D.M. Couillard, Z.J. Stansell, M.W. Farnham, Genetic diversity and population structure of collard landraces and their relationship to other *Brassica oleracea* crops, Plant Genome 8 (3) (2015) 1–11, https://doi.org/10.3835/plantgenome2015.04.0023.
[46] F.G. Maraş-vanlıoğlu, H. Yaman, F. Kayaçetin, Genetic diversity analysis of some species in Brassicaceae family with ISSR markers, Biotech Studies 29 (1) (2020) 38–46, https://doi.org/10.38042/biost.2020.29.01.05.
[47] A. Okumus, A. Balkaya, Estimation of genetic diversity among Turkish kale populations (*Brassica oleracea* var. acephala L.) using RAPD markers, Russ. J. Genet. 43 (4) (2007) 411–415, https://doi.org/10.1134/S1022795407040096.
[48] K.H. Singh, L. Singh, N. Parmar, S. Kumar, J. Nanjundan, G. Singh, A.K. Thakur, Molecular characterization and genetic diversity analysis in Indian mustard (*Brassica juncea* L. Czern & Coss.) varieties using SSR markers, PLoS One 17 (8) (2022) e0272914, https://doi.org/10.1371/journal.pone.0272914.
[49] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909, https://doi.org/10.1038/ng1847.
[50] J. Yu, G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, E.S. Buckler, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2006) 203–208, https://doi.org/10.1038/ng1702.
[51] J. Zou, C. Jiang, Z. Cao, R. Li, Y. Long, S. Chen, J. Meng, Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping, Genome 53 (2010) 908–916, https://doi.org/10.1139/G10-075.
[52] W. Ecke, R. Clemens, N. Honsdorf, H.C. Becker, Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.), Theor. Appl. Genet. 120 (5) (2010) 921–931.
[53] J. Xu, N. Ranc, S. Munos, S. Rolland, J.P. Bouchet, N. Desplat, M.C. Le Paslier, Y. Liang, D. Brunel, M. Causse, Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species, Theor. Appl. Genet. 126 (2013) 567–581, https://doi.org/10.1007/s00122-012-2002-8.