BMC Plant Biology

RESEARCH Open Access

# The genetics of fruit skin separation in date palm

Shameem Younuskunju[1,5], Yasmin A. Mohamoud[1], Lisa Sara Mathew[4], Klaus F. X. Mayer[5,6], Karsten Suhre[3] and Joel A. Malek[1,2*]

**Abstract**

The physical appearance of date palm (Phoenix dactylifera) fruit (dates) is important for its market value. Many date-producing countries experience significant financial losses due to the poor appearance of the fruit, skin separation or puffiness being a major reason. Previous research showed evidence linking the skin separation phenotype to environmental conditions. To investigate this further, a genome-wide association study was conducted using genome data from 199 samples collected from 14 countries. Here, we identified nine genetic loci associated with this phenotype and investigated genes in these regions that may contribute to the phenotype overall. Multiple genes in the associated regions have functional responses to growth regulators and are involved in cell wall development and modification. Analysis of gene expression data shows many are expressed during fruit development. We show that there are both environmental and genetic contributions to the fruit skin separation phenotype. Our results indicate that different date cultivars exhibit varying degrees of skin separation despite genetic similarities or differences. However, genetically different cultivars show extreme differences compared to the phenotype variation between genetically similar cultivars. We demonstrate that beyond environmental factors, genetics is a strong contributor to the most extreme skin separation in some cultivars. Identifying the genetic factors may help better understand the biology and pathways that lead to the environmental effects on skin separation and improve commercial date production. In conclusion, our key finding is that both environmental and genetic factors contribute to skin separation variation, and improvements in environmental factors alone cannot overcome the extreme level of variation observed in some cultivars.

**Keywords** Significant SNP, Fruit skin, GWAS, Genetic region, Puffiness, Microcracks, Fruit skin separation, Genetic variation, Association study

*Correspondence:
Joel A. Malek
jom2042@qatar-med.cornell.edu
[1]Genomics Laboratory, Weill Cornell Medicine-Qatar, Doha 24144, Qatar
[2]Department of Genetic Medicine, Weill Cornell Medicine-Qatar, Doha 24144, Qatar
[3]Department of Physiology, Weill Cornell Medicine-Qatar, Doha 24144, Qatar
[4]Clinical Genomics Laboratory, Sidra Medicine, P.O Box 26999, Doha, Qatar
[5]Shool of Life Sciences, Technical University of Munich, 85354 Munich, Germany
[6]Plant Genome and Systems Biology, Helmholtz Center Munich, 85764 Munich, Germany

## Introduction

Fruit quality is an essential factor in the market value of commercial agriculture. Biochemical composition and physical features are key parameters in determining quality, and the appearance of the fruit is an important consideration for consumers [1–3]. The visual appeal of date palm fruits (dates) is adversely affected by exocarp separation from the mesocarp (commonly called skin separation or puffiness) and microcracks, which usually occur during the ripening stage [4–6]. Every year, growers experience significant financial loss because of fruit skin appearance [6, 7]. Skin separation phenomena affect the shelf-life of dates and customers' purchasing choices [8]. Understanding and addressing this phenotype in commercially important fruits, including dates, is an ongoing topic of research [9, 10].

Date palm (Phoenix dactylifera L.) is an economically important crop in the Arabian Peninsula, North Africa and Pakistan, and plays a significant role in the economic development of these regions [11, 12]. Thousands of date palm varieties grow in hot, arid habitats worldwide; each variety exhibits a wide range of fruit characteristics, including sugar content, moisture, size and colour [13–15]. Dates are a rich source of sugar, phenolic antioxidants, fiber, and proteins, making them economically significant worldwide [12, 16]. The uniformity of colour and size, sugar content, and the absence of visual defects are some of the criteria for grading the quality of dates for marketing [11, 13, 15, 17]. It is hypothesized that many of these economically significant phenotypes are linked to genetic features, and multiple recent studies have investigated this possibility, including our own association study showing genetic control of dry fruit colour (Tamar stage) [14, 18–21].

To take advantage of the long shelf life of dates requires the maintenance of undamaged skin throughout pre and post-harvest periods [8, 22]. The fruit skin (exocarp) is made up of cuticle, epidermis, and hypodermis and is considered an essential element in fleshy fruits [10, 23, 24]. It provides mechanical protection from biotic and abiotic stresses as well as contributes to visual appearance [4, 25]. Skin separation phenomenon does not occur in all date varieties and is mainly observed in economically significant varieties like Barhi, Sagai, Sukari, Khalas, Kheneizi, and Medjool [6, 9, 24, 26]. Many studies have focused on understanding the factors causing this phenotypic variation, ranging from microclimatic and nutritional aspects to the mechanical characteristics of the date's cell wall [6, 7, 9, 24, 26]. A physiochemical study of Sukari dates showed that environmental conditions, irrigation, and fertilisation methods are factors that can improve this and other fruit traits [26]. Variations in the mechanical behaviour of different date cultivars, such as Dayri (no skin separation) and Barhi (extreme variation of skin separation), have been associated with skin separation phenomena [24]. Furthermore, the study of Medjool dates showed that environmental factors and cyclic stresses of turgor pressure fluctuations can also influence the traits [9]. Another study suggested that climate factors are not the only contributors to this phenotype by showing that the percentage of sclereid cells was significantly higher in skin-separated than in normal fruit [6].

As we have observed, most previous studies examined the influence of environmental factors on skin separation in dates. However, we hypothesized that genetic factors may also play a part in determining which date cultivars are most affected by this phenotype. Evidence from other plants, such as research on bell peppers, has shown that genetic variation may impact skin separation [27, 28] and encouraged this line of investigation. To our knowledge, no studies have been conducted to date to understand the influence of genetic factors on this disorder in date palm, likely because the impact of the environment is clear within a specific cultivar. To address that, the dataset used in this study is extensively diverse in origin and variety [29], which is critical for understanding genetic associations within the context of a phenotype also affected by the environment. Importantly, because we collected the same cultivars from multiple environmental locations, our unique dataset could help distinguish the range of environmental effects versus genetic effects on the skin separation phenotype.

## Materials and methods

### Phenotypic data

Digital photographs of dry fruits from each cultivar were used to measure the phenotypic data. The fruits from different cultivars were collected across multiple years (2012 to 2015) but not from the same trees. The same cultivar might have been collected in some years, but those collections were from different countries. Each cultivar had 5 to 11 fruits from the same collection as representatives (Supplementary file: Figure. S1). We manually assessed multiple images of each cultivar and scored the skin separation variation. The score was rated from 0 to 10 based on skin defects. Fruit with no defects was rated with a score of 0, and 10 was the maximum score for complete defects. A total of 1637 images belonging to 171 cultivars were manually assessed, and the skin defect rate was scored for each image. Outliers were removed from the raw dataset using the Z-score method ( with ±2 standard deviations). One researcher conducted multiple rounds of scoring for digital images of fruit. In each round, the scoring of individual fruits exhibited minor variations compared to previous scores due to the subjective nature of the analysis. However, the skin separation score for multiple fruits within each cultivar were assigned independently, and the mean score for each

cultivar was subsequently calculated. This mean score effectively adjusted for the variance observed across multiple rounds for each cultivar. The average skin separation rate was calculated using the outlier-removed raw data for each sample. We then performed a BoxCox transformation on the average score to reduce skewness. Box cox-transformed data were used as phenotypic data for the genome-wide association study.

#### Genome sequencing and SNP calling

We used a genome dataset of date samples from our previous association study of fruit colour [14]. Sequencing libraries were constructed from total DNA extracted from fruits, and whole genome libraries were sequenced using Illumina 2500/4000 instruments. For a more detailed description of the sequencing data, please refer to the studies of Mathew et al. and Thareja et al. [20, 30]. The quality control processing of samples (QC), raw reads, genome alignment, and SNP calling was carried out in accordance with our previous association study on Tamar stage date fruit colour [14]. SNPs were marked as missing if DP < 10 and filtered with the following parameters using VCftools (V0.1.16), genotype call rate 80%, minor allele frequency 0.01, and Hardy-Weinberg equilibrium $1 \times 10^{-6}$.

#### Genetic similarity and phenotypic variation analysis

Genetic relatedness between the samples (kinship coefficients) was measured using Plink software (v1.9) [31] with the 'make-king-table' option. The resulting data were filtered using a kinship score greater than or equal to 0.354 to find genetically similar samples. Genetically similar samples were grouped based on pairwise kinship scores and considered for the phenotype variation analysis of genetically identical and dissimilar cultivars if a group had at least three or more samples from different regions. Skin separation variation analysis of samples within a group was performed to assess phenotypic differences between genetically similar cultivars grown in different regions and environments. To compare the phenotypic variation between genetically different cultivars, we assessed the differences among samples from different groups.

#### Genome-wide association study

Genome-wide analysis (GWAS) was performed using the FarmCPU method [32] implemented in the GAPIT (v3) R package [32]. LD pruning was performed on the QC-filtered SNP dataset using the Plink software (--indep-pairwise 500 50 0.99) to improve the computational efficiency of the GWAS method [31]. A kinship matrix and four principal components (PCA) were used as covariates in the GWAS to correct for population structure. Both the PCA and kinship matrix scores were calculated from the LD-pruned SNPs using the GAPIT R package. The VanRaden algorithm in the GAPIT R package was used to measure the kinship matrix. A list of significant SNPs associated with phenotype was identified using a cutoff value of FDR-adjusted *p*-values of 0.05 (5%). Manhattan and QQ plots were generated from association results using the CMPlot R package [33].

#### Structural variation and RNA-seq analysis of candidate genes

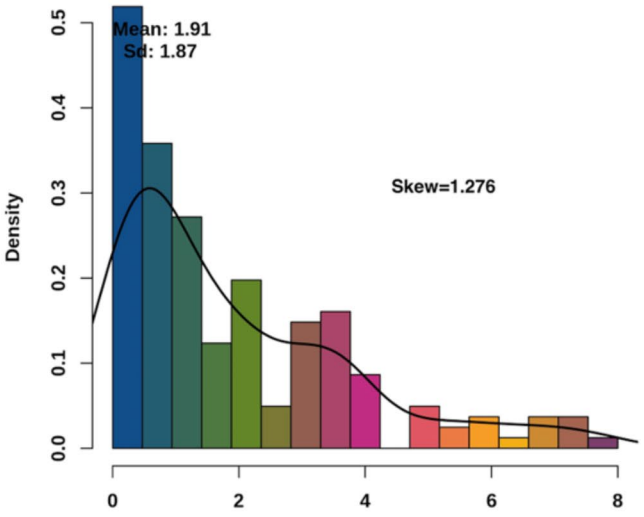
Regions spanning 100 kb upstream and downstream of GWAS-identified significant SNPs were examined to determine potential candidate genes and variants. The gene sequences were obtained from these regions using the GFF3 annotation file of the PDK50 reference genome (PRJNA40349). Gene Ontology analysis of the candidate gene sequences was conducted using Blast2GO software [34]. Gene functions were determined through literature reviews and Blast2Go results. All INDELs and SNPs from the regions were annotated using SNPEff software [35]. The SNPEff annotated SNPs and INDELs from the potential regions were filtered with an LD R2 value >= 0.6 to the significantly associated SNPs. RNA-seq expression analysis was carried out using the transcriptome data of kheneizi and Khalas cultivars from Hazzouri et al.'s study [36]. The data contain three or four replicates taken at different post-pollination days (DPP) in two cultivars (45,75,105,120,135 days). Read alignment and expression analysis was performed as described in our previous association study on fruit colour [14]. Structural variation (SV) analysis was conducted by utilising clipped, discordant, unmapped, and indel reads from each sample that was homozygous for the alternative allele with respect to the reference genome, as these would contain SVs not already observed in the reference. See Younuskunju et al. [14] for details on the analysis of structural variations.

## Results

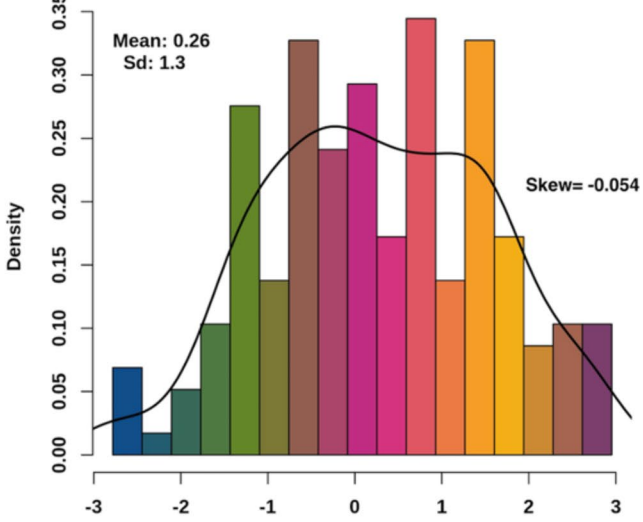
#### Phenotypic data analysis: skin separation rate

Manual analysis of the fruit images showed that the rate of skin separation varied from cultivar to cultivar (Fig. 1). The averaged dataset showed that 62 samples had a skin separation score greater than or equal to 2, while 109 samples had a score less than 2. The minimum score observed was 0, and the maximum score reached 8. The distribution analysis of phenotype indicated a positive skew (right) with a skewness value of 1.27 (Fig. 1a). To improve the data quality and reduce skewness for the association study, a BoxCox transformation was carried out [37]. This transformation reduced the skewness of the phenotypic data to -0.054 (Fig. 1b). During the transformation, scores less than or equal to 0.99 were converted to negative values, and scores greater than or equal to 1 were converted to positive values (Fig. 1c). This BoxCox

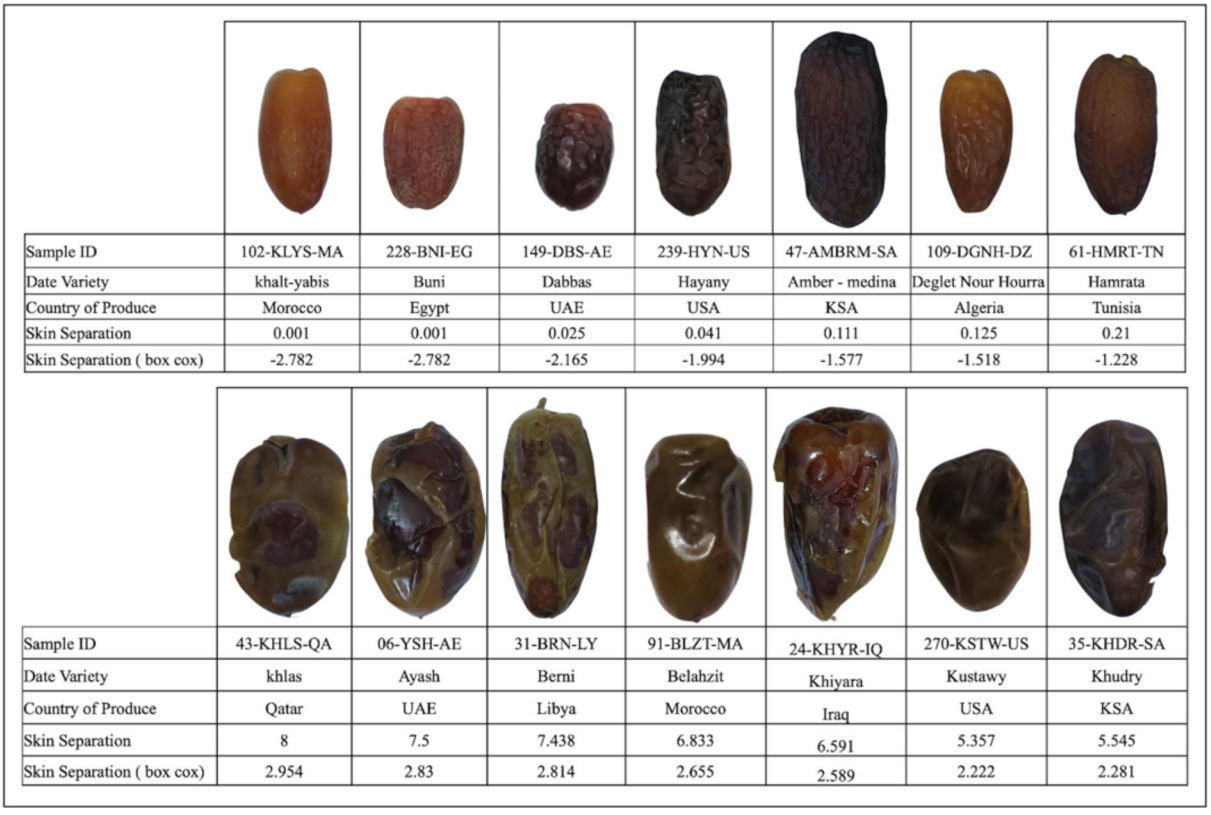
(a)




(b)




(c)



| Sample ID | 102-KLYS-MA | 228-BNI-EG | 149-DBS-AE | 239-HYN-US | 47-AMBRM-SA | 109-DGNH-DZ | 61-HMRT-TN |
|---|---|---|---|---|---|---|---|
| Date Variety | khalt-yabis | Buni | Dabbas | Hayany | Amber - medina | Deglet Nour Hourra | Hamrata |
| Country of Produce | Morocco | Egypt | UAE | USA | KSA | Algeria | Tunisia |
| Skin Separation | 0.001 | 0.001 | 0.025 | 0.041 | 0.111 | 0.125 | 0.21 |
| Skin Separation ( box cox) | -2.782 | -2.782 | -2.165 | -1.994 | -1.577 | -1.518 | -1.228 |

| Sample ID | 43-KHLS-QA | 06-YSH-AE | 31-BRN-LY | 91-BLZT-MA | 24-KHYR-IQ | 270-KSTW-US | 35-KHDR-SA |
|---|---|---|---|---|---|---|---|
| Date Variety | khlas | Ayash | Berni | Belahzit | Khiyara | Kustawy | Khudry |
| Country of Produce | Qatar | UAE | Libya | Morocco | Iraq | USA | KSA |
| Skin Separation | 8 | 7.5 | 7.438 | 6.833 | 6.591 | 5.357 | 5.545 |
| Skin Separation ( box cox) | 2.954 | 2.83 | 2.814 | 2.655 | 2.589 | 2.222 | 2.281 |

**Fig. 1** Distribution and comparison analysis of fruit skin separation phenotype values. We manually assessed multiple fruit images of each cultivar and scored skin separation ranging from 0 to 10. The distribution analysis showed that the phenotype data were positively skewed (right) with a skewness of 1.27, so we performed a Box-Cox transformation to reduce the skewness of the phenotypic data. (**a**) distribution analysis plot of raw phenotype values. (**b**) distribution plot of box-cox transformed phenotype values. (**c**) representative image of 7 date cultivars with the lowest and highest skin separation scores

transformed dataset was used as the phenotypic data for the association study (Additional file 1).

### Genotyping and SNP calling

Quality control (QC) filters of the raw genotypic dataset produced 188 samples and 10,183,993 SNPs across the 18 linkage groups (LGs) and unplaced scaffolds in the reference genome. Haplotype linkage analysis showed that genotype correlation (R2) decreased to half its maximum at 25.9 kb (half LD decay value). For more detailed results of SNP calling and LD decay analysis, please refer to the study by Younuskunju et al. [14]. LD pruning resulted in 3.541 million SNPs that were utilized in the association study (Additional file 2).

### Genetic relatedness and phenotypic variation analysis

Genetic relatedness analysis resulted in the identification of sample pairs with high kinship coefficient scores (0.354) suggesting effectively the same genotype (Supplementary file: Figure. S2). These genetically similar samples fell into 23 different groups. To compare the differences in skin separation among genetically similar cultivars grown in different locations and environments, we chose five sample groups out of the 23, each containing at least three or more samples from different regions (Supplementary file: Table S1). These groups are named Deglet Nour, Medjoul, Mabroom, Safawi, and Sagai. The analysis of phenotypic variation was carried out within each group separately, and then averages were compared between groups (Figs. 2 and 3). The samples in the Deglet Nour group exhibited scores ranging from 0.12 to 1.1 with average 0.58, Medjoul from 1 to 2 with average 1.76, Mabroom from 2 to 3.63 with average 2.75, Safwai from 1.49 to 2.72 with average 2.06, and Sagai from 3.9 to 5.04 with average 4.46 (Fig. 2). The phenotype comparison between different cultivar groups (inter-cultivar) showed higher phenotypic differences than did the intra-cultivar comparison (Fig. 3). The comparison between the Deglet Nour, Medjoul & Sagai groups showed significant differences (Wilcoxon test $p$-value) (Fig. 3b). The comparison of the Deglet Nour group to the Sagai group showed the most extreme difference in skin separation between the groups ($p$-value=0.0061).

### Association of fruit skin separation and significant SNP genotypic effects

Genome-wide association using the box-cox transformed phenotypic data resulted in the discovery of several significant SNPs. The QQ plot of the association showed a lambda score of 0.97, indicating that the test statistics aligned with the expected distribution (Fig. 4a). The FDR-adjusted $p$-value cutoff of 5% ( FDR<0.05) identified nine SNPs that were significantly associated with the phenotype (Table 1; Fig. 4b and Additional file 3). These SNPs were located in multiple linkage groups (LG) in the PDK50 reference genome. Among the identified significant SNPs, LG8s1288981, LG3s2983485, LG6s6604213, LG18s8763935, LG10s3647074, and LG4s35891773 SNPs showed significant Wilcoxon test $p$-values (Fig. 5 and Supplementary file: Figure. S3). Analysis of the 3 possible genotypes at the associated SNPs showed the effect of each allele (Fig. 5). Further

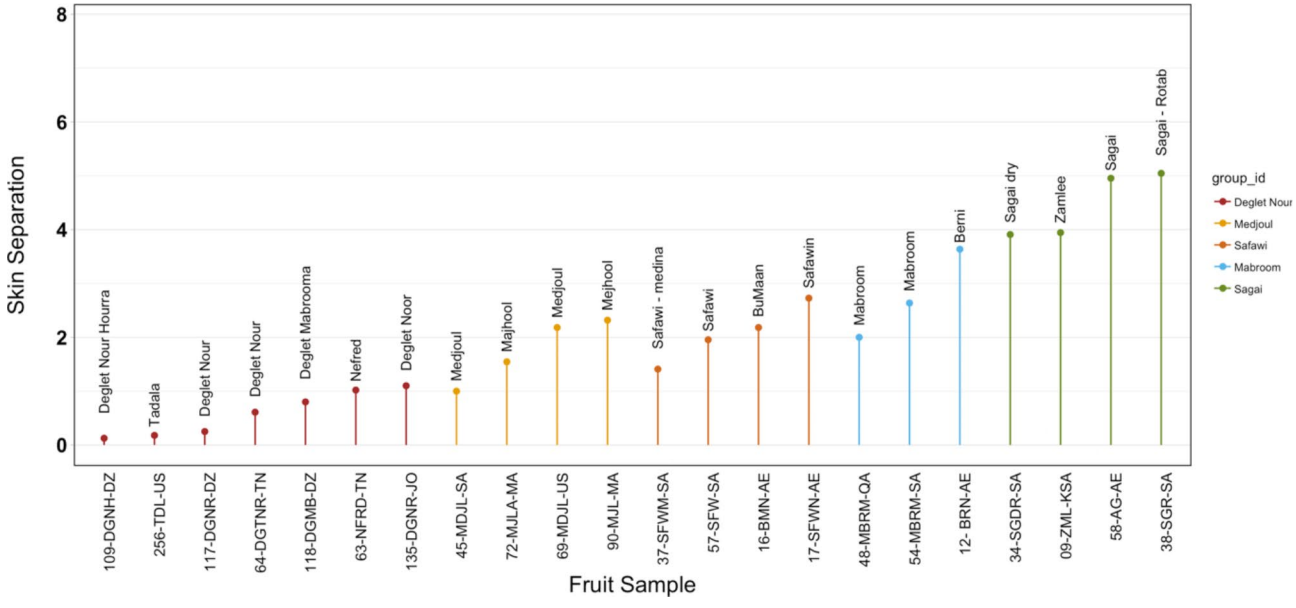



**Fig. 2** Comparison analysis of the skin separation variation of genetically similar fruit cultivars grown in different regions and environments. Genetically similar cultivars were marked with a separate colour code. The X-axis represents samples from multiple cultivars, and the Y-axis represents the skin separation rate

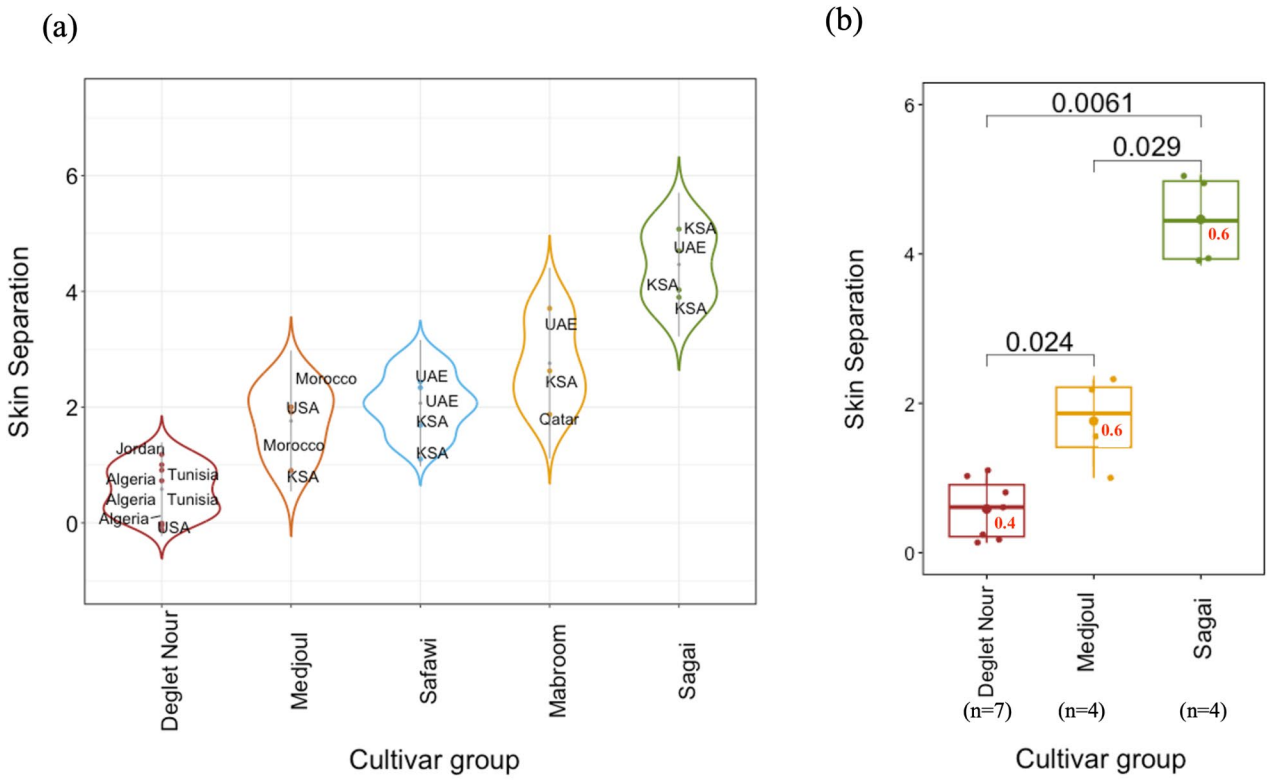



**Fig. 3** Comparison analysis of fruit skin separation of genetically different fruit cultivars. The cultivar groups named Deglet Nour, Medjoul, Mabroom, Safawi, and Sagai were chosen for the analysis (**a**): Skin separation comparison of five genetically different cultivars. Each point represents a sample and is marked as the country of origin. (**b**): boxplot distribution and standard deviation analysis of skin separation in three genetically different cultivar groups. *P*-values were calculated using the Wilcoxon statistical test. The standard deviation of skin separation scores for fruit samples within each cultivar group is marked in red colour. The X-axis represents the cultivar group, and the Y-axis represents the skin separation rate

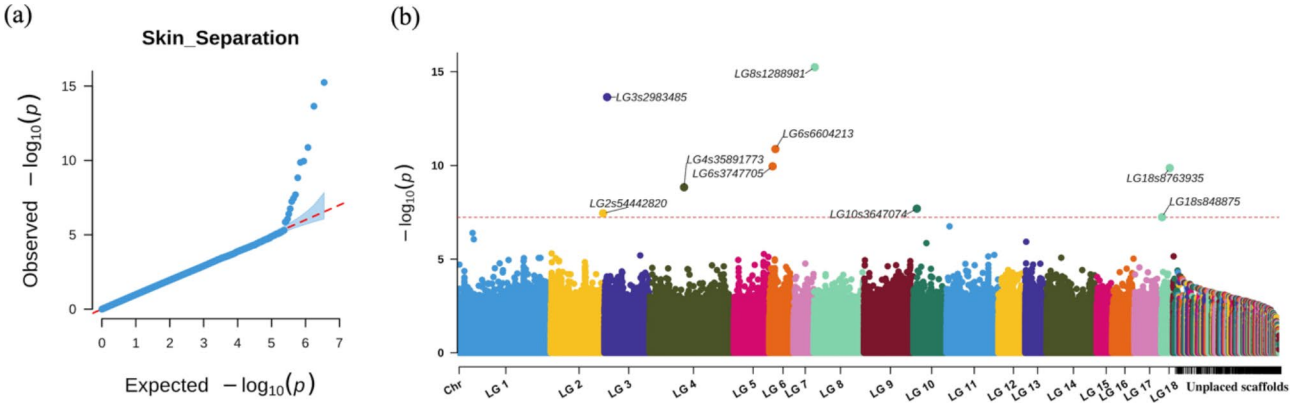



**Fig. 4** Genome-wide association study (GWAS) analysis of skin separation phenotype using the LD pruned SNP set of 3.541 million SNPs.(**b**, **c**) QQ plot and Manhattan plot using the LD pruned SNP set for all Linkage group and unplaced scaffolds

validation was provided by the visual examination of the phenotype when separated by the genotypes of SNP LG18s8763935 (Fig. 6). It was observed that the fruit exhibited a higher level of skin separation rate when the sample was homozygous for the reference allele (REF) at SNP LG18s8763935, whereas the fruit showed a lower rate when the sample was homozygous for the alternative allele (ALT) at SNP LG18s8763935.

## Candidate gene and SNP annotation

A total of 169 genes were identified across the associated regions (Additional file 4). The results of the literature search and Blast2GO analysis showed that many of these genes are involved in lignin synthesis, plant-type cell wall loosening, cell wall organisation, and response to auxin, abscisic acid, and gibberellic acid (Table 2). To identify potential causal genes for skin separation

**Table 1** List of significant SNPs associated with skin separation phenotype from association study (GWAS). A false-discovery rate (FDR) adjusted *p*-value was used as a cut-off for identifying significant SNPs associated with the phenotype

| SNP | LG ID | MAF | *P*.value | FDR *P*.Value |
|---|---|---|---|---|
| LG8s1288981 | LG8 | 0.14 | 5.71E-16 | 2.02E-09 |
| LG3s2983485 | LG3 | 0.21 | 2.26E-14 | 4.00E-08 |
| LG6s6604213 | LG6 | 0.07 | 1.34E-11 | 1.58E-05 |
| LG6s3747705 | LG6 | 0.05 | 1.11E-10 | 9.64E-05 |
| LG18s8763935 | LG18 | 0.46 | 1.36E-10 | 9.64E-05 |
| LG4s35891773 | LG4 | 0.05 | 1.46E-09 | 0.00086216 |
| LG10s3647074 | LG10 | 0.35 | 2.01E-08 | 0.01017247 |
| LG2s54442820 | LG2 | 0.09 | 3.57E-08 | 0.01581493 |
| LG18s848875 | LG18 | 0.22 | 5.76E-08 | 0.02267 |

expressed in fruit within these candidate regions, we analyzed gene expression data. RNA-seq data from the Khalas and Kheneizi varieties showed that many of these genes were indeed expressed during various stages of fruit development (postpollination days) (Fig. 7). Among them, Expansin, Cellulose synthase-like protein, Myb transcription factor, Ras-related protein Rab-8 A, C2 and GRAM domain-containing protein, Soluble inorganic pyrophosphatase and Transport inhibitor response 1-like protein genes were expressed during the early stages of development (dpp 45–70) and the Proteasome subunit beta type, DOF zinc finger protein1 were expressed during the later stages (dpp 105–135). The SNPEff annotated SNPs and INDELs from the candidate genes with an LD R2 value>=0.6 to the significantly associated SNPs showed many putative high and moderate effects on the encoded proteins (Additional file 5). Structural variation (SV) analysis was conducted for all potential regions of the significantly associated SNPs as described in the association study of date fruit colour. The SV pipeline revealed that the potential regions of SNPs LG3s2983485 and LG10s3647074 exhibited multiple SVs, including insertion, deletion and repeat expansion (Additional file 6). Despite finding structural variations in these regions, none of them demonstrated a strong correlation with the genotypic variation of significant SNPs identified from the GWAS result.

## Discussion

Previous studies on date skin separation phenomena have demonstrated strong evidence that environmental factors play a significant role [6, 9, 24, 26, 38]. Determining the genetic association with a phenotype in the presence of environmental factors presents a considerable challenge [39, 40]. The study requires vast and diverse sample datasets from various environments. In this study, we used a unique dataset to determine the effect of genetic factors on skin separation phenomena in the presence of background environmental effects. We collected a variety of genetically similar cultivars from different environments and locations. These genetically similar but environmentally distinct samples enable us to begin to determine the extent to which the environment versus genetics influences the trait. The analysis showed a range of phenotypic variations within genetically similar samples (intra-cultivar analysis) and between cultivars (inter-cultivar analysis).

The analysis of genetically similar cultivars grown in different regions showed a range of skin separation between the samples of the same cultivar (Fig. 2) based on environment, though presumably this range is related to commercial growth conditions only and not more extreme conditions such as drought. That is, the phenotypic range in these cultivars is due to differences in factors such as watering, post-harvest treatment and other abiotic effects. We observe significant differences highlighting the importance of environmental factors on skin separation, as observed in past studies. However, importantly, the inter-cultivar analysis demonstrated a higher level of variation compared to the intra-cultivar analysis of samples grown in distinct regions and environments (Figs. 2 and 3). That is, the variation due to environment appears less than that due to genetics when the

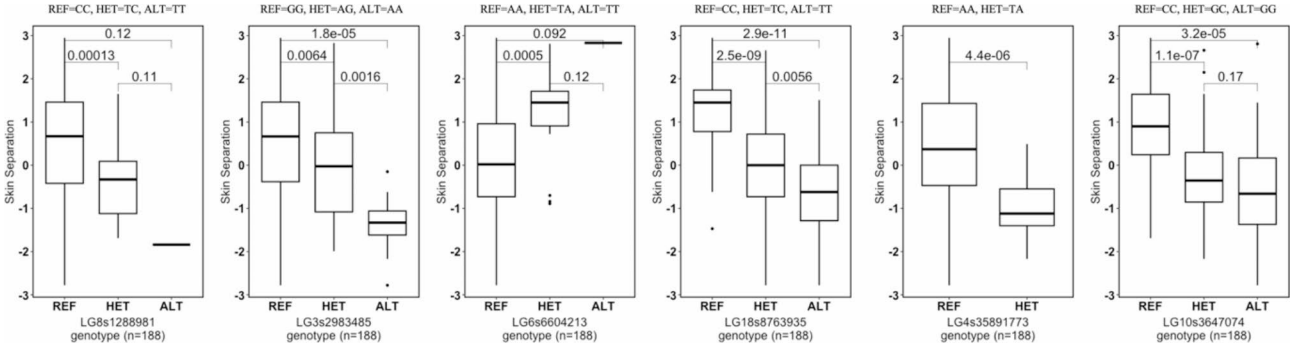



**Fig. 5** Boxplot distribution analysis of skin separation phenotypes by genotypes of a list of significant SNPs from this association study (GWAS). These five SNPs have significant Wilcoxon test *p*-values among the list of GWAS-identified SNPs. The *X*-axis represents the SNPs' genotypes, and the *Y*-axis represents the phenotypic value (box-cox transformed phenotype)

(a)

| Sample ID | 43-KHLS-QA | 06-YSH-AE | 31-BRN-LY | 08-NWDR-AE | 24-KHYR-IQ | 15-JBR-AE | 23-MLMD-IQ |
|---|---|---|---|---|---|---|---|
| Date Variety | Khlas | Ayash | Berni | Nawader | Khiyara | Jabri | Um Alamod |
| LG18s8763935 - Genotype | REF | REF | REF | REF | REF | REF | REF |
| Skin Separation | 8 | 7.5 | 7.4375 | 6.8 | 6.59090909 | 6.04545455 | 6 |
| Skin Separation ( box cox) | 2.95409342 | 2.82983354 | 2.81390878 | 2.64602391 | 2.58863541 | 2.43285598 | 2.41944937 |

(b)

| Sample ID | 87-TZGRT-MA | 102-KLYS-MA | 239-HYN-US | 233-NGL-AE | 109-DGNH-DZ | 106-TFNY-DZ | 80-AGA-MA |
|---|---|---|---|---|---|---|---|
| Date Variety | Tazgaghzt | Khalt-yabis | Hayany | Naghal | Deglet Nour Hourra | Teeflahny | Agalid |
| LG18s8763935 - Genotype | ALT | ALT | ALT | ALT | ALT | ALT | ALT |
| Skin Separation | 0.001 | 0.001 | 0.04166667 | 0.0875 | 0.125 | 0.15625 | 0.15909091 |
| Skin Separation ( box cox) | -2.7823506 | -2.7823506 | -1.9947294 | -1.6918469 | -1.5185721 | -1.399669 | -1.3896921 |

**Fig. 6** Differences in fruit skin separation when samples were categorised by the genotypes of LG18s8763935 SNP. Fruits were grouped together which are homozygous for the (**a**) RFE or (**b**) ALT allele of SNP LG18s9876335. Results show that fruit skin separation is at an extreme level when the sample is homozygous for the REF allele and at a very low level or none when the sample is homozygous for the ALT allele at SNP LG18s8763935

cultivars with the highest and lowest skin separation phenotype are compared. This significant variation (Fig. 3b) between genetically different cultivars, even in the presence of variation from the environment, suggests that genetic factors are the primary contributors to extreme differences in this phenotype. That is, genetic and environmental factors contribute to the differences in skin separation phenotype, but the genetic factor is the strongest contributor to extreme differences. These findings, particularly under varying environmental conditions, underscore the complex interaction between genetics and the environment.

The GWAS results showed nine markers associated with the phenotype (Fig. 5 and Supplementary file: Figure. S3). The outcomes of the analysis for SNPs LG8s1288981, LG3s2983485, LG18s8763935, LG4s35891773, and LG10s3647074 demonstrate that the skin separation rate is considerably higher in the cultivar when the sample is homozygous for the reference allele, and the rate decreases or is absent when the sample is homozygous ALT or HET with respect to the reference allele. Visual examination of the phenotype when separated by the genotypes of SNP LG18s8763935 (Fig. 6) supports these observations. Studies have shown that changes in the biochemical properties of exocarps during fruit ripening can lead to skin separation and microcracking [8, 41]. Identifying the genes responsible for the development and modification of cell walls and cuticular membranes will lead to a better understanding of the contribution of genetic factors to these skin disorders. Our study identified several key genes involved in cell wall development and modification in the regions surrounding significantly associated SNPs (Table 2). For example, the Expansin gene is in proximity to the LG6s6604213 SNP. This gene plays a crucial role in cell wall loosening and weakening during cell expansion [42, 43]. Cosgrove's study showed that plant cells produce expansin protein during growth, which unlocks the polysaccharide wall network and allows turgor-induced cell walls to loosen [42]. The region surrounding LG3s2983485 contains the Soluble inorganic pyrophosphatase gene, which is involved in cell wall thickening and the metabolic process of cell wall pectin [44]. Furthermore, the candidate regions from the LG3s2983485, LG4s35891773, LG8s1288981, and LG18s8763935 SNPs contain many genes, including Myb transcription factor, DOF zinc finger protein,

**Table 2** List of genes detected around the 100 kb region of significant SNPs from association study (GWAS) result associated with the skin separation phenotype. Genes were selected if they have a putatively significant role in fruit growth regulation, lignin synthesis, and plant-type cell wall development

| Tag SNP | Protein Name | Gene function |
|---|---|---|
| LG18s8763935 | Proteasome subunit beta type | response to water deprivation; response to ethylene; response to abscisic acid; brassinosteroid mediated signaling pathway; positive regulation of auxin mediated signaling pathway; |
| LG3s2983485 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein | response to gibberellin; cellular response to water deprivation; cellular response to abscisic acid stimulus |
| LG3s2983485 | Soluble inorganic pyrophosphatase | cell wall thickening; cell wall pectin metabolic process |
| LG3s2983485 | Ras-related protein Rab-8 A | cell wall biogenesis |
| LG4s35891773 | C2 and GRAM domain-containing protein | positive regulation of abscisic acid-activated signaling pathway |
| LG6s6604213 | Expansin | response to gibberellin; unidimensional cell growth; plant-type cell wall loosening; structural constituent of cell wall |
| LG6s6604213 | Myb transcription factor | response to water deprivation; response to ethylene; response to abscisic acid; organ boundary specification between lateral organs and the meristem; positive regulation of auxin mediated signaling pathway |
| LG6s6604213 | DOF zinc finger protein 1 | response to auxin; response to salicylic acid; cell wall modification; positive regulation of cell cycle |
| LG6s6604213 | Cellulose synthase like | cell wall organization |
| LG6s6604213 | Cellulose synthase-like protein | cell wall organization |
| LG8s1288981 | Cyclin-D4-1 | cell division |
| LG8s1288981 | Cyclin-D3-1 | response to cytokinin; response to sucrose; guard mother cell differentiation; cell division |
| LG8s1288981 | DNA-directed RNA polymerase D subunit 1-like protein | regulation of cell division |
| LG8s1288981 | Transport inhibitor response 1-like protein | auxin-activated signaling pathway; response to jasmonic acid |

Proteasome subunit beta and basic helix-loop-helix DNA-binding protein (bHLH), all of which respond to growth regulators such as Auxin, Gibberellic acid (GA), and abscisic acid [42, 45–48]. Studies on apples and lychee have shown that growth regulators play an important role in skin cracks during fruit development [4, 49]. Specifically, research on apples has found that increased epidermal density is associated with greater resistance to cracking, and GA may enhance epidermal cell density [49, 50]. Xiangdong Fu et al. study (2002) showed that Proteasome-mediated proteolysis plays a significant role in the degradation of DELLA proteins, a key step in the GA signalling pathway [51]. Our gene expression results revealed that Proteasome subunit beta type and DOF zinc finger protein1 are expressed at higher levels during the later stages of fruit development, particularly at the Rutab stage (dpp-135), while other genes were expressed predominantly at earlier stages (Fig. 7). The later stage gene expression of these two genes is interesting given that a study of the date palm fruit by Alsmairat and colleagues showed anatomical differences linked to the skin separation phenomenon occurring in the later stages of fruit development, especially in the Rutab stages [6]. Additionally, a study in Arabidopsis revealed that overexpression of the vascular-related DOF1 (VDOF1) gene increases lignin deposition (lignification) [52]. This higher lignin level makes the cell wall more rigid and less elastic, which can contribute to the formation of skin cracks [53]. These insights suggest that genes from our candidate region, particularly the proteasome subunit beta type and DOF zinc finger protein1, may be important in the skin separation phenotype. Their expression patterns align with the timing of observed anatomical changes and the effects of growth regulators, however further research is needed to assign any potential functional role in the phenotype.

Our study findings help reveal the various contributions of environmental and genetic factors to skin-separation phenotypic variation in dates. While our study highlights the contributions of both genetic and environmental factors, one key limitation of our study is the lack of environmental variables, such as watering, heat, sunlight exposure, and post-harvest treatment conditions, which are known to influence the skin separation phenotype. Future studies that document this information would allow a more comprehensive and controlled Genotype x Environment (GxE) study analysis and provide clearer insights into how the contribution to phenotype. We confirm that environmental factors likely modify skin separation given intra-cultivar variation. However, in some cultivars like Sagai, the genetic factors are so significant that environmental improvement may only result in minor effects improvements when compared to skin separation in other cultivars. We recognize that another limitation in the study is that the samples represent 199 cultivars collected from 14 countries, mostly grown

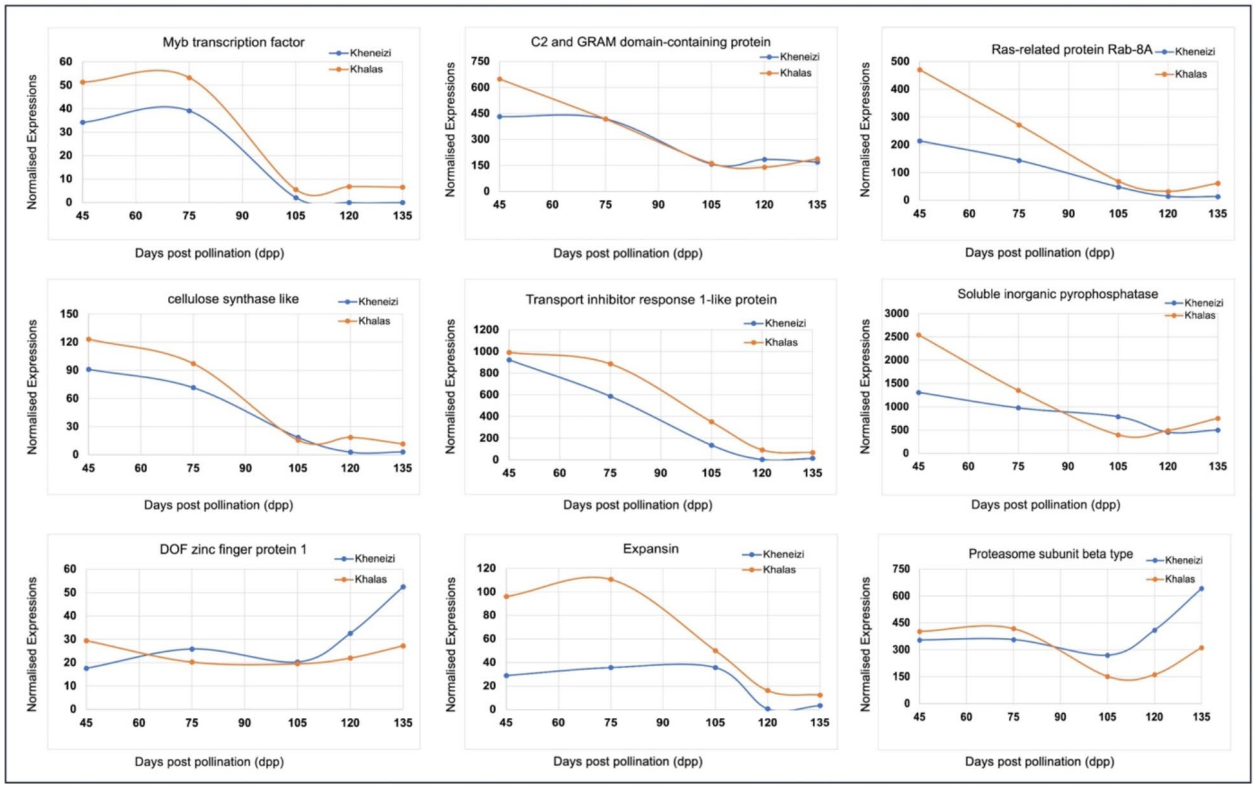



**Fig. 7** RNA seq analysis of expression of genes identified in the potential region around identified SNPs. The gene expression analysis was carried out across three or four replicates of fruit development stages of two varieties, namely Kheneizi (red) and Khalas (yellow). Each point on the X-axis represents the development stage (post-pollination date), while the Y-axis displays the mean normalised expression read count across three or more replicates

under commercial production conditions. This sample size may not adequately capture the genetic diversity or environmental variability that affects skin separation across all date cultivars. A more extensive dataset with larger sample sizes and broader environmental representation beyond the commercial growth conditions that we collected from will be necessary for future research to enhance our understanding of the true range of phenotype and genetic contribution. Our study contributes to understanding the influence of environmental and genetic factors on skin separation in the most popular date palm cultivars. We expect this knowledge will contribute to the process of selecting and developing fruit varieties with reduced skin separation in future breeding programs.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05713-1.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

**Acknowledgements**
We thank Robert Krueger of the USDA, Sean Lahmeyer of Huntington Gardens and Diego, Rivera of the University of Murcia for contributing date palm, Phoenix and palm species.

**Author contributions**
JAM envisioned the project, designed the study, analysed data, and wrote & reviewed the main manuscript; SY designed the study, analysed data, conducted bioinformatics analysis, prepared figures, and wrote the main manuscript; LSM conducted genome sequencing; YAM directed library construction and sequencing; KS designed the study and reviewed the manuscript. KFXM designed the study and reviewed the manuscript.

**Funding**
This study was made possible by grant NPRP-EP X-014-4-001 from the Qatar National Research Fund (a member of Qatar Foundation). The funding agency did not participate in the study design, sample collection, analysis, data interpretation or writing of this research.

**Data availability**
WGS data of 188 samples used in this study are being submitted to NCBI SRA, which is in progress. Table 1 includes the SNP association results of date palm skin separation phenotypes on the date palm reference genome. Gene annotation from the associated linkage group is included in Table 2.

## Declarations

**Ethics approval and consent to participate**
Human and animal subjects were not included in this study.

**Consent for publication**
Not Applicable.

**Competing interests**
The authors declare no competing interests.

Received: 4 July 2024 / Accepted: 15 October 2024
Published online: 06 November 2024

**References**

1. Alam MZ, et al. Contributing factors to quality of date (Phoenix dactylifera L.) fruit. Sci Hort. 2023;321:112256.
2. Mikulic-Petkovsek M, et al. Fruit Quality characteristics and biochemical composition of fully ripe blackberries harvested at different Times. Foods. 2021;10(7):1581.
3. Barrett DM, Beaulieu JC, Shewfelt R. Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: desirable levels, instrumental and sensory measurement, and the effects of processing. Crit Rev Food Sci Nutr. 2010;50(5):369–89.
4. Khadivi-Khub A. Physiological and genetic factors influencing fruit cracking. Acta Physiol Plant. 2014;37(1):1718.
5. Santos M et al. Molecular mechanisms involved in fruit cracking: a review. Front Plant Sci. 2023;14.
6. Alsmairat N, et al. Anatomical Assessment of skin separation in date palm (Phoenix dactylifera L. var. Mejhoul) Fruit during Maturation and ripening stages. Agriculture. 2023;13:38.
7. Lobo MG, Yahia EM, Kader AA. Biology and Postharvest Physiology of Date Fruit, in Dates. 2013;57–80.
8. Lara I, Heredia A, Domínguez E. Shelf Life potential and the Fruit cuticle: the unexpected player. Front Plant Sci, 2019. 10.
9. Lustig I, Bernstein Z, Gophen M. Skin separation in Majhul fruits. Int J Plant Res. 2014;4(1):29–35.
10. Fernández-Muñoz R, Heredia A, Domínguez E. The role of cuticle in fruit shelf-life. Curr Opin Biotechnol. 2022;78:102802.
11. Chao CT, Krueger RR. The date palm (Phoenix dactylifera L.): overview of Biology, uses, and cultivation. HortScience Horts. 2007;42(5):1077–82.
12. Al-Shahib W, Marshall RJ. The fruit of the date palm: its possible use as the best food for the future? Int J Food Sci Nutr. 2003;54(4):247–59.
13. Ahmed S, et al. Phenotypic diversity of date palm cultivars (Phoenix dactylifera L.) from Algeria. Gayana Botánica. 2016;73:42–53.
14. Younuskunju S, et al. Genome-wide association of dry (Tamar) date palm fruit color. Plant Genome. 2023;16(4):e20373.
15. Elhoumaizi MA, et al. Phenotypic diversity of date-palm cultivars(Phoenix dactylifera L.) from Morocco. Volume 49. Genetic Resources and Crop Evolution; 2002. pp. 483–90. 5.
16. Ghnimi S, et al. Date fruit (Phoenix dactylifera L.): an underutilized food seeking industrial valorization. NFS J. 2017;6:1–10.
17. Zaid A, Arias-Jimenez EdJ. Date palm cultivation. FAO Plant Production and Protection Paper (FAO); 1999.
18. Hazzouri KM, et al. Genome-wide association mapping of date palm fruit traits. Nat Commun. 2019;10(1):4680.
19. Malek JA, et al. Deletion of beta-fructofuranosidase (invertase) genes is associated with sucrose content in date palm fruit. Plant Direct. 2020;4(5):e00214.
20. Mathew LS, et al. A genome-wide survey of date Palm cultivars supports two major subpopulations in Phoenix dactylifera. G3 (Bethesda). 2015;5(7):1429–38.
21. Al-Dous EK, et al. De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). Nat Biotechnol. 2011;29(6):521–7.
22. Ahmed ZFR, et al. Improving Fruit Quality, Bioactive compounds, and Storage Life of date palm (Phoenix dactylifera L., Cv. Barhi) using natural elicitors. Horticulturae. 2021;7(9):293.
23. Kayesh E et al. Fruit skin color and the role of anthocyanin. Acta Physiol Plant, 2013. 35.
24. Gophen M. Skin separation in date fruits. Int J Plant Res. 2014;4:11–6.
25. Ginzberg I, Stern RA. Strengthening fruit-skin resistance to growth strain by application of plant growth regulators. Sci Hort. 2016;198:150–3.
26. Ghazzawy HS, et al. Combined impact of Irrigation, Potassium Fertilizer, and thinning treatments on yield, skin separation, and Physicochemical properties of date palm fruits. Plants. 2023;12(5):1003.
27. Kline WL, et al. Evaluation of Six Nitrogen Fertility Programs on Marketable Yield and Development of skin separation in Bell Pepper Fruit. HortTechnology Hortte. 2011;21(3):323–8.
28. Wyenandt CA, et al. Production system and Cultivar effects on the development of skin separation or silvering. Bell Pepper Fruit HortTechnology Hortte. 2017;27(1):37–44.
29. Stephan N, et al. A comprehensive metabolomic data set of date palm fruit. Data Brief. 2018;18:1313–21.
30. Thareja G, et al. Genotyping-by-sequencing identifies date palm clone preference in agronomics of the state of Qatar. PLoS ONE. 2018;13(12):e0207299.
31. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
32. Wang J, Zhang Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Proteom Bioinf. 2021;19(4):629–40. Genomics.
33. Yin L, et al. rMVP: a Memory-efficient, Visualization-enhanced, and parallel-accelerated Tool for Genome-wide Association study. Volume 19. Genomics, Proteomics & Bioinformatics; 2021. pp. 619–28. 4.
34. Conesa A. S. Götz 2008 Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008 p619832.
35. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.
36. Hazzouri KM, et al. Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. Nat Commun. 2015;6:8824.
37. Cieleń G, et al. The impact of Box-Cox transformation on phenotypic and genomic characteristics of litter size variability in Landrace pigs. Animal. 2023;17(5):100784.
38. Bitar AaA-Q. Hassan Effect of Water Stress at Fruit Maturity Stage on production and skin separation phenomenon of date palm cv. Medjool. Hebron Univ Res Journal-A (Natural Sciences). 2021;10(1):Article 1.
39. Wojcik GL, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514–8.
40. Aschard H, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. Hum Genet. 2012;131(10):1591–613.
41. Petit J, et al. Analyses of tomato fruit brightness mutants uncover both cutin-deficient and cutin-abundant mutants and a new hypomorphic allele of GDSL lipase. Plant Physiol. 2014;164(2):888–906.
42. Cosgrove DJ. Loosening of plant cell walls by expansins. Nature. 2000;407(6802):321–6.
43. Brummell DA, et al. Modification of Expansin Protein Abundance in Tomato Fruit alters Softening and Cell Wall Polymer Metabolism during Ripening. Plant Cell. 1999;11(11):2203–16.
44. Li N, et al. Overexpression of UDP-glucose pyrophosphorylase from Larix gmelinii enhances vegetative growth in transgenic Arabidopsis thaliana. Plant Cell Rep. 2014;33(5):779–91.
45. Zou X, Sun H. DOF transcription factors: specific regulators of plant biological processes. Front Plant Sci, 2023. 14.
46. Santner A, Estelle M. The ubiquitin-proteasome system regulates plant hormone signaling. Plant J. 2010;61(6):1029–40.
47. Marowa P, Ding A, Kong Y. Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep. 2016;35(5):949–65.
48. Hao Y et al. Basic Helix-Loop-Helix (bHLH) transcription factors regulate a wide range of functions in Arabidopsis. Int J Mol Sci, 2021. 22(13).
49. Joshi M, et al. Identification of candidate genes mediating apple fruit-cracking resistance following the application of gibberellic acids 4 + 7 and the cytokinin 6-benzyladenine. Plant Physiol Biochem. 2018;127:436–45.
50. Eccher T, Hajnajari H. Fluctuations of endogenous gibberellin A4 and A7 content in apple fruits with different sensitivity to russet in X International Symposium on Plant Bioregulators in Fruit Production 727. 2005.
51. Fu X, et al. Gibberellin-mediated proteasome-dependent degradation of the barley DELLA protein SLN1 repressor. Plant Cell. 2002;14(12):3191–200.
52. Ramachandran V, et al. Plant-specific Dof transcription factors VASCULAR-RELATED DOF1 and VASCULAR-RELATED DOF2 regulate vascular cell differentiation and lignin biosynthesis in Arabidopsis. Plant Mol Biol. 2020;104(3):263–81.
53. Grimm E, et al. Structural and physiological changes associated with the skin spot disorder in apple. Postharvest Biol Technol. 2012;64(1):111–8.

## Publisher's note