

Re-sequencing and transcriptomic analysis reveals rich DNA sequence variation and differential gene expression between varieties of *Ziziphus jujuba*

Na Li

Shanxi Agricultural University

Yuqin Song

Shanxi Agricultural University

Jie Li

Shanxi Agricultural University

Ruijie Hao

Shanxi Agricultural University

Xinxin Feng

Shanxi Agricultural University

Liulin Li (✉ tgliulin@163.com)

Shanxi Agricultural University

Research Article

Keywords: *Ziziphus jujuba*, transcriptome, genome re-sequencing, indel, nitrite reductase

Posted Date: March 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-253917/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Jujube is one of the characteristic fruit tree species in China. 'Linhuang No. 1', a cracking-resistant cultivar, and 'Muzao', a cracking-susceptible cultivar, were selected as materials by previous study. Whole-genome re-sequencing and transcriptome of 'Linhuang No. 1' and 'Muzao' allow the screening out of differently expressed genes with different gene structures between them. It could be helpful in explaining divergence/similarity of cracking resistance between the two cultivars. **Results:** There are 664,129 mutation sites between 'Linhuang No. 1' and 'Muzao' by re-sequencing. To determine the genetic relationship of 'Linhuang 1', 'Muzao' and reference genome 'Dongzao', the characteristic mutation sites were analyzed by principal component analysis. The genetic relationship between 'Linhuang No. 1' and 'Muzao' was closer than that with 'Dongzao'. A total of 431 differentially expressed genes was screened by transcriptomics, and 19 differentially expressed genes were screened by combining the transcriptomics with re-sequencing analysis. LOC107427052 (encoding nitrite reductase) was determined by KEGG enrichment analysis for further study. **Conclusions:** The large base insertion was not in the domain region of the LOC107427052 gene CDS region. As verified by the finding that the base insertion did not affect protein translation. Our study has laid a foundation for the analysis of genetic information and the comparative nitrite metabolism of 'Linhuang No. 1' and 'Muzao'.

Background

Jujube is one of the characteristic fruit tree species in China[1]. Jujube is valued as a traditional herbal medicine, and is cultivated on 2 million hectares in China alone, with an annual production of approximately 7.36 million tons in 2018 (<https://data.stats.gov.cn/easyquery.htm?cn=E0103>). Jujube cv. Muzao is one of the most widely cultivated jujube varieties in China, combining the characteristics of high yield and high quality[2]. Due to the wide cultivation area of the Muzao tree crop (*Ziziphus jujube* Mill.), there are many varieties and geographical types of jujube, varying in terms of fruit shape, single fruit weight, drying rate, nutrient composition and fruit characteristics[3]. 'Linhuang No. 1' is a dry-jujube cultivar selected from 'Muzao', which has fruit characteristics similar to those of 'Muzao'. Compared with 'Muzao', 'Linhuang No. 1' shows greater resistance to fruit cracking[4]. 'Linhuang No. 1' has become a valuable genotype for scientific research.

Whole-genome resequencing is the process of sequencing the genomes of different individuals of a species where the genome has already been sequenced[5]. Through sequence alignment, this methodology is able to identify large numbers of single nucleotide polymorphisms (SNPs), insertion/deletion (indel) variants, structural gene variants (SV), copy number variants (CNV) and other mutations in the whole-genome re-sequenced individuals. This information on genetic variation is an important way to understand the genetic background of a species and to study its evolution. In order to explore the wide phenotypic diversity of sweet cherry varieties with respect to important agronomic traits, such as flowering time and defense response to pathogens, Aliko et al. (2020) re-sequenced 21 cherry accessions and found that the majority of high-impact SNPs (e.g., addition of stop codon, frameshifts) were identified in genes involved in flowering time, dormancy and defense reactions against pathogens[6]. Yu et al. (2018) re-sequenced the genomes of 58 peach cultivars and closely related species to explore the origin and evolutionary history of peaches[7]. The results showed that peaches originated on the Qinghai-Tibet Plateau in Southwest China and had been subjected to frugivore-mediated selection to drive the evolution of fruit traits. The publication of the 'Dongzao' genome sequence provided valuable help for genomic and transcriptomic studies of other jujube cultivars[8]. The genome sequence of jujube cv. Dongzao represents a valuable resource not only for biological discovery and crop improvement but also for evolutionary and comparative genomic analysis. Huang et al. (2016) re-sequenced the genomes of 31 cultivated and wild jujubes to reveal the domestication process of jujube and screened key genes for metabolic effects associated with sweet or sour fruit[9].

In this article, we present analyses of whole-genome re-sequencing and transcriptome of 'Linhuang No. 1' and 'Muzao'. The sequence analysis focused on genomic regions associated with propitious variation, such as single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels). The analysis focused on genes differentially expressed during the full-red period of fruit development. The combination of whole-genome re-sequencing and transcriptome analysis would allow the screening out of differently expressed genes with different gene structures between 'Linhuang No. 1' and 'Muzao', which could be helpful in explaining divergence/similarity of cracking resistance between the two cultivars.

Results

Discovery of SNPs and InDels in 'Linhuang No.1' and 'Muzao'

A total of 110 million raw reads were generated for each of 'Linhuang No. 1' and 'Muzao'. After removal of low-quality reads, about 95.93% ('Linhuang No. 1') and 94.74% ('Muzao') of the reads were retained as clean data and used for further investigation. The Q20 percentage (proportion of nucleotides with a quality value greater than 20 in the reads) was greater than 95% and the Q30 percentage was greater than 89%. The mean GC percentage of the clean reads was 33%. The high-quality reads were further mapped onto the reference genome (*Ziziphus jujube* 1.1), using BWA software. Overall, almost 94% of these reads were uniquely mapped, and covered about 46.5% of the reference genome, with at least 30 × coverage depth (Supplementary materials 1).

Additional file 1: Table S1. Summary of sequencing and mapping results

Compared with the *Z. jujuba* reference ('Dongzao') genome sequence, a total of 4,404,985 polymorphic sites including 3,623,127 SNPs and 781,858 indels were discovered in 'Linhuang No. 1', with a total of 4,401,491 polymorphic sites, including 3,626,419 SNPs and 775,072 indels, in 'Muzao'. From the comparison between 'Linhuang No. 1' and 'Muzao', 664,129 polymorphisms were detected. Table 1(A) summarizes the number of SNP and indel variants relative to the reference genome sequence and according to chromosome, whereas Table 1(B) details the number of polymorphisms between 'Linhuang No. 1' and 'Muzao' by chromosome. The highest number of SNPs and indels between 'Linhuang No. 1' and the reference genome was detected for NC_029679.1 (Chromosome 1), whereas the lowest number of SNPs and indels was detected for NC_029690.1 (Chromosome 12). For all chromosomes, indel polymorphisms were less frequent than SNPs. From the comparison between 'Muzao' and the reference genome, the highest and lowest number of SNPs and

indels were also detected on Chromosomes 1 and 12, respectively. From the comparison between 'Linhuang No. 1' and 'Muzao', the highest and lowest numbers of SNP and indel polymorphisms were also found on Chromosomes 1 and 12, respectively.

Fig. 1 shows the relatedness among 'Linhuang No.1', 'Muzao' and 'Dongzao'. The distance between the points represents the degree of characteristic differences among the accessions. The first and the second principal components (PC1 and PC2) explained 58.22% and 31.92% of the total variation, respectively. The cumulative contribution rate of the two principal components was 90.14%. PC1 differentiated 'Linhuang No. 1' and 'Muzao' from 'Dongzao', whereas PC2 differentiated 'Linhuang No. 1', 'Muzao' and 'Dongzao'. From the comparison between 'Linhuang No. 1' and 'Muzao', 664,129 polymorphisms were detected.

SNP and indel density plots, obtained through the comparison of 'Linhuang No. 1' and 'Muzao' against the reference genome 'Dongzao', are displayed in Fig. 2. This figure also provides an integrated view of the polymorphisms detected across the entire genome between 'Linhuang No. 1' and 'Muzao' in relation to the reference genome. 'Linhuang No. 1' has a similar level of polymorphism to 'Muzao', when compared with the reference genome.

The changes in gene expression patterns detected by RNA-seq

To further investigate the gene expression patterns of 'Linhuang No. 1' and 'Muzao', transcriptome sequencing was used. Genes showing more than two-fold up- or down-regulation between the two jujube cultivars were classified as "differentially expressed genes" (DEGs). In total, 'Linhuang No. 1' and 'Muzao' were covered by an average of 27.55 and 28.21 million clean reads, respectively (Supplementary Fig. 1). Supplementary Fig 2 shows the genes with common expression and cultivar-specific expression levels between 'Linhuang No. 1' and 'Muzao' (expression threshold=1). A total of 13,150 genes were co-expressed in 'Linhuang No. 1' and 'Muzao' and a total of 702 and 1,188 genes were expressed in only 'Linhuang No. 1' or 'Muzao', respectively (Supplementary Fig. 2).

Additional file 2: Figure S1. Statistics of sequencing data from jujube cultivars Linhuang No.1 and Muzao

Additional file 3: Figure S2. Venn map of cultivar specificity of expression of individual genes

Cluster Analysis of DEGs

Data and analyses of DEGs show that there were 431 significant DEGs in 'Linhuang No. 1' and 'Muzao' (Fold Change=2). Of the 431 DEGs, 25 genes were expressed only in 'Linhuang No. 1', while 11 genes were expressed only in 'Muzao' (Supplementary materials 2 the special expression gene in LH and M). We performed cluster analysis on DEGs identified in the 'Linhuang No. 1' and 'Muzao' comparisons, and found that, of the 431 DEGs, 303 genes were significantly up-regulated in 'Muzao' relative to 'Linhuang No.1' (Supplementary Fig. 3a) and 128 genes were down-regulated in 'Muzao' (Supplementary. Fig 3b). We analyzed clustering of the log₁₀ (FPKM) of the samples' gene expression. In Supplementary Fig. 3ab, a color closer to red represents greater gene expression. In Supplementary Fig. 3c, each column represents 'Linhuang No.1' versus 'Muzao' and each row represents a DEG presented in different colors representing fold change, increased (red) or decreased (blue).

Additional file 4: Table S2. The specific expressed genes in 'Linhuang No.1' and 'Muzao'

Additional file 5: Figure S3. Analysis of differentially expressed genes (DEGs). A: up-regulated DEGs in 'Muzao' relative to 'Linhuang No.1'; B: down-regulated DEGs in 'Muzao' relative to 'Linhuang No.1'

GO Analysis

A total of 431 DEGs was analyzed by GO classification. The results were divided into three categories, namely the biological process, the cellular component, and the molecular function. (Fig. 3A) shows GO functional distribution of the DEGs from the 'Linhuang No. 1' and 'Muzao' comparison. The top five high-enrichment GO terms from the 'Linhuang No. 1' and 'Muzao' comparison were metabolic process, catalytic activity, cellular process, single-organism process and binding, with 162, 141, 140, 118 and 115 genes being annotated into each term, respectively.

Enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis

The statistics of the top 20 significantly overrepresented pathways of differentially expressed genes (DEGs) from 'Linhuang No. 1' vs 'Muzao' are shown in Fig. 3B. Among these KEGG pathways, carbon metabolism and flavonoid biosynthesis were both enriched with respect to 11 genes. According to the significance standard of $P < 0.05$, there were ten significantly enriched pathways in the comparison Table 2.

Table 2. Significantly enriched KEGG Pathways in 'Linhuang No. 1' compared with 'Muzao'

pathway ID	pathway	'Linhuang No.1' VS 'Muzao' (p value)
1	ko00941	Flavonoid biosynthesis 3.65E-09
2	ko00061	Fatty acid biosynthesis 0.000819
3	ko00640	Propanoate metabolism 0.001013
4	ko00620	Pyruvate metabolism 0.001408
5	ko01212	Fatty acid metabolism 0.004836
6	ko00780	Biotin metabolism 0.009491
7	ko01200	Carbon metabolism 0.009727
8	ko04712	Circadian rhythm - plant 0.016141
9	ko00410	beta-Alanine metabolism 0.021581
10	ko00562	Inositol phosphate metabolism 0.047753

Differential expression of different structural genes between 'Linhuang No. 1' and 'Muzao'

In order to further explore the differential expression of variant genes between 'Linhuang No. 1' and 'Muzao', combined resequencing and transcriptome analysis of 'Linhuang No. 1' and 'Muzao' was carried out. The results showed that there were 33 and 29 SNP and indel variant genes, respectively, differentially expressed in 'Linhuang No. 1' and 'Muzao'. A total of 947 genes in 'Linhuang No. 1' and 'Muzao' were mutated by SNP or indel, and 19 genes were differentially expressed between the two cultivars (Fig 4.A).

We performed cluster analysis on the 19 DEGs (Fig 4.B). We analyzed clustering of the log₂ (Fragments per Kilobase Million, FPKM) of the gene expression. A color closer to red represents higher gene expression. Of the 19 DEGs, 12 genes were up-regulated in 'Muzao', including LOC107413306, LOC107427052, LOC107431206, LOC107432789, LOC107427373, LOC107404912, LOC107428467, LOC107426536, LOC107414473, LOC107434555, LOC107405940, and LOC107405347.

The 19 differentially expressed genes were further enriched by KEGG analysis (Fig 4. C, Table 3). According to the standard of $p < 0.05$, there were three enriched pathways in 'Linhuang No. 1' and 'Muzao', namely stilbenoid, diarylheptanoid and gingerol biosynthesis (Rich factor: 14.98), flavonoid biosynthesis (Rich factor: 8.85) and nitrogen metabolism (Rich factor: 19.48). Nitrogen metabolism (LOC107427052) was selected to further verify the transcriptomic data.

Table 3 Enriched pathways of 19 differentially expressed genes

pathway ID	pathway	p-value	
1	ko00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	0.00697
2	ko00941	Flavonoid biosynthesis	0.01965
3	ko00910	Nitrogen metabolism	0.04939
4	ko00450	Selenocompound metabolism	0.06990
5	ko00940	Phenylpropanoid biosynthesis	0.07128
6	ko09106	Metabolism of other amino acids	0.10862
7	ko00270	Cysteine and methionine metabolism	0.13537
8	ko04626	Plant-pathogen interaction	0.17866
9	ko09159	Environmental adaptation	0.20606
10	ko09150	Organismal Systems	0.20606
11	ko09110	Biosynthesis of other secondary metabolites	0.26609
12	ko00460	Cyanoamino acid metabolism	0.26972
13	ko04144	Endocytosis	0.38498
14	ko09102	Energy metabolism	0.38498
15	ko04075	Plant hormone signal transduction	0.40467
16	ko09140	Cellular Processes	0.49486
17	ko09141	Transport and catabolism	0.49486
18	ko04131	Membrane trafficking	0.52736
19	ko09132	Signal transduction	0.56775
20	ko09105	Amino acid metabolism	0.56775
21	ko09100	Metabolism	0.63880
22	ko09130	Environmental Information Processing	0.67476
23	ko09182	Protein families: genetic information processing	0.97181
24	ko09180	Brite Hierarchies	0.99815

Levels of nitrite reductase (NiR) activity, nitrite and ammonia nitrogen concentration

LOC107427052 encodes nitrite reductase. In the process of assimilating nitrate nitrogen in plants, nitrite reductase (NiR) is coupled with nitrate reductase (NR) to complete the inorganic assimilation of nitrate. Nitrite reductase can catalytically reduce nitrite to ammonia. Nitrite reductase plays a very important role in the growth and development of plants and is widespread in the leaves and roots of higher plants. Based on the re-sequencing of the genomes of 'Linhuang No. 1' and 'Muzao', it was found that indel events occurred in the LOC107427052 sequences of both 'Linhuang No. 1' and 'Muzao', compared with 'Dongzao' (Fig 5). Some 106 bases were inserted at position 12517784 (in the coding sequence (CDS) region) on chromosome NC_029687.1 of 'Linhuang No. 1', while 104 bases were inserted at the same position of 'Muzao'. In order to confirm the results of re-sequencing, the LOC107427052 gene was cloned, using cDNA of 'Linhuang No. 1' and 'Muzao'. The protein sequences encoded by the LOC107427052 gene of 'Dongzao', 'Linhuang No. 1' and 'Muzao' were compared, and the similarity was 91.82% (Supplementary Fig. 4). Domain prediction was performed on the protein sequence, and it was found that 'Linhuang No. 1' and 'Muzao' base insertion did not occur in the domain region (Supplementary Fig. 5). Real-time PCR was used to verify the expression level of the LOC107427052 gene (Fig. 6). The results showed that the LOC107427052 gene expression levels of 'Muzao' were significantly higher than those of 'Linhuang No. 1' during fruit development, and that the expression levels of the LOC107427052 gene in 'Linhuang No. 1' and 'Muzao' increased gradually as the fruit developed. The expression level of the LOC107427052 gene was significantly higher during the coloring period and the full-red period than in the young fruit period.

Additional file 6: Figure S4. Sequence alignment of 'Dongzao', 'Linhuang No. 1' and 'Muzao' LOC107427052 gene

Additional file 7: Figure S5. Domain prediction of 'Dongzao', 'Linhuang No. 1' and 'Muzao' LOC107427052 gene

At the same time, the nitrite reductase activities of 'Linhuang No. 1' and 'Muzao' at different developmental stages were determined. The results showed that there was no significant difference in the nitrite reductase activities between 'Linhuang No. 1' and 'Muzao' at the later stage of fruit development; The nitrite reductase activities of 'Linhuang No. 1' and 'Muzao' were 0.75 U/g fresh weight (FW) and 0.93 U/g FW at the full-red stage, respectively. There was no significant difference in the concentrations of the substrate nitrite and the product ammonia of nitrite reductase between the two varieties.

Discussion

The Chinese jujube tree was domesticated from the wild jujube (*Z. jujuba* Mill. var. *spinosa* Hu.) [10]. Cultivated jujube trees and wild jujube showed different characteristics, such as tree vs. shrub habit, sparsely thorned vs. heavily thorned, and large vs. small fruits, respectively, through artificial selection for important agronomic characters [8, 11, 12]. Most cultivated jujube varieties produce relatively fewer seeds due to self-incompatibility or cross-incompatibility. Jujube trees can only be bred through seed progeny selection, selection by bud mutation or molecular breeding [13]. These genetic methods resulted in the high heterozygosity, high repeat sequence density and low GC content of the jujube genome [13]. Therefore, sequencing of the jujube genome was very difficult. Liu et al. (2014) published the genome of the first cultivar of the genus *Ziziphus*, 'Dongzao' (437.65 Mbp) [8]. The release of this genome sequence provided a rich resource of genetic information for the breeding of improved jujube accessions, as well as for the molecular improvement of other plants and fruit trees of the family Rhamnaceae.

Huang et al. (2016) reported the genome sequence of another jujube variety, 'Junzao', and re-sequenced the genomes of 31 wild jujube and jujube accessions with different geographical distributions [9]. The research of Huang and colleagues revealed the genomic mechanism underlying improvement of fruit sweetness and acidity during domestication and identified four genes related to acidic metabolism pathways, encoding NADP-dependent malic enzyme, pyruvate kinase, isocitrate dehydrogenase and aconitate hydratase, which play key roles in organic acid metabolism. This research provides valuable genomic information and material for jujube breeding. Guo et al. (2020) reported the re-sequencing of 350 jujube accessions [13]. Through a genome-wide association study (GWAS) and selective sweep analysis, variation in the genes involved in the regulation of seven domestication traits were screened for, including fruit shape, kernel size and fruiting branch length. This study provided rich genomic resources for revealing the genetic basis of the domestication and evolution of jujube.

In the current study, the varieties 'Linhuang No. 1' (resistant to cracking) and 'Muzao' (susceptible to cracking) were used as the research materials, with 'Dongzao' used as the reference genome for resequencing analysis. It was found that 4,404,985 and 4,401,491 polymorphic sites were obtained in 'Linhuang No. 1' and 'Muzao', respectively, whereas 664,129 variant sites were found between 'Linhuang No. 1' and 'Muzao'. Principal component analysis on the characteristic variant sites between 'Linhuang No. 1', 'Muzao' and 'Dongzao' were performed and the results showed that both PC1 and PC2 could distinguish 'Linhuang No. 1' and 'Muzao' from 'Dongzao', indicating that 'Linhuang No. 1' and 'Muzao' were closer in relation to one another than to 'Dongzao'.

Mutation is a key element in species evolution. The widespread base substitution and insertion/deletion mutations are important driving forces for genome evolution [14, 15]. Insertion and deletion mutations are more likely to trigger species evolution than are base mutations. The greater the evolutionary distance between species, the more bases are inserted or deleted, and the greater the length of the insertions or deletions [16-18]. Insertions and deletions are the main reasons for the divergence of closely related species. The insertion and deletion of bases can cause DNA sequence changes and cause DNA fragment length polymorphisms. They can even change the structure of genes through insertions or deletions in the exons and introns of the original gene, leading to the generation of new genes in the genome [19]. It was found, through re-sequencing, that there were 781,858 and 775,072 base insertions or deletions in 'Linhuang No. 1' and 'Muzao', respectively. Compared with 'Linhuang No. 1', there were 169,267 base insertions and deletions in 'Muzao'. To a large extent, this led to the different characteristics between 'Linhuang No. 1' and 'Muzao'.

The transcriptome combines the genetic information of the genome with that of the proteome, with biological functions based on RNA, uses high-throughput sequencing technology (RNA-Seq) to sequence all cDNA libraries in tissues or cells and calculates the gene expression under different processing conditions (by counting the number of relevant reads). The jujube transcriptome has been widely used in research into response to heat stress, cold stress, salt stress and coloring of jujube fruits [20-22]. In the current study, 431 differentially expressed genes were identified by transcriptome analysis of 'Linhuang No. 1' and 'Muzao' pericarps. At the same time, the transcriptome and resequencing analyses were combined to screen for differential expression of 'Linhuang No. 1' and 'Muzao' genes with respect to gene structural mutations. A total of 19 mutant genes were screened for differential expression. There were three significantly enriched pathways in 'Linhuang No. 1' and 'Muzao': stilbenoid, diarylheptanoid and gingerol biosynthesis (Rich factor: 14.98), flavonoid biosynthesis (Rich Factor: 8.85) and nitrogen metabolism (Rich factor: 19.48). The gene LOC107427052, encoding nitrite reductase, was screened for further study.

In the process of nitrate assimilation in plants, nitrite reductase (NiR) is coupled with nitrate reductase (NR) to complete the inorganic assimilation of nitrate. Nitrite reductase can catalytically reduce nitrate to ammonium [23]. Ozawa and Kawahigashi (2005) cloned the NiR gene of rice 'Konansou' and overexpressed the gene in a commercial rice variety, 'Koshihikari' [24]. The results showed that, compared with the wild type 'Koshihikari', the introduced exogenous NiR gene made the plant grow better, while callus regeneration ability was also significantly improved.

Compared with the reference genome 'Dongzao', 'Linhuang No. 1' had an insert of 106 bases at the position 12517784 (CDS region) on chromosome NC_029681.1, whereas 'Muzao' had an insert of 104 bases at the same position; fragment insertion was verified by cloning. Insertion/deletion mutations are rarer than base substitutions in the evolution of organisms. If the sequence of the coding region has an indel mutation involving a base sequence which is not a multiple of three bases, it will cause serious consequences as a result of a frameshift mutation, which places the mutant under greater pressure in natural selection. In the present study, the domain of the protein encoded by the cloned NiR sequence was predicted, and it was found that the base insertion did not occur in the domain region. The expression levels of the LOC107427052 genes of 'Linhuang No. 1' and 'Muzao' were consistent with the changes measured in nitrite reductase activity during fruit development. However, the LOC107427052 gene expression level of 'Muzao' was significantly higher than that of 'Linhuang No. 1' during the coloring period and the full-red period of fruit development, but there was no significant difference in nitrite reductase activity. There was also no significant difference in the concentrations of the substrate nitrite and the product ammonia of nitrite reductase between the two varieties.

Conclusions

This present study was the first to explore the differences between different jujube varieties (Linhuang No. 1 and Muzao) by combining genome resequencing and transcriptomics. It found, by re-sequencing, that there are 664,129 mutation sites between 'Linhuang No. 1' and 'Muzao'. The characteristic mutation sites between 'Linhuang No. 1', 'Muzao' and 'Dongzao' were analyzed by principal component analysis. The genetic relationship between 'Linhuang No. 1' and

'Muzao' was closer than that with 'Dongzao'. A total of 431 differentially expressed genes was screened by transcriptomics, and 19 differentially expressed genes were screened by combining the transcriptomics with re-sequencing analysis. LOC107427052 (encoding nitrite reductase) was determined by KEGG enrichment analysis for further study. There was a large base insertion in the CDS region of the LOC107427052 gene, but the insertion was not in the domain region, as verified by the finding that the base insertion did not affect protein translation. Our study has laid a foundation for the analysis of genetic information and the comparative nitrite metabolism of 'Linhuang No. 1' and 'Muzao'.

Methods

Plant materials

Fruits and leaves of 'Linhuang No. 1' and 'Muzao' were harvested from trees from the Jujube Germplasm Resource Nursery, Pomology Institute, Shanxi Agricultural University, Shanxi, China. The trees were 50 years old and cultivated according to current regulations for integrated fruit production. Fruits and leaves without disease or physical injuries were selected for sampling. All fruits and leaves were packed and delivered immediately to the laboratory at Shanxi Agricultural University.

Genome re-sequencing

Young leaves of 'Linhuang No. 1' and 'Muzao' were collected and stored at -80°C prior to DNA isolation. Genomic DNA was extracted using a modified CTAB method. The sequencing library was prepared according to the standard protocol of Illumina and sequencing was conducted by Lianchuan bio, Beijing, China, using a Illumina NovaSeq 6000. Clean data were aligned to the jujube reference genome (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/826/755/GCF_000826755.1_ZizJuj_1.1/GCF_000826755.1_ZizJuj_1.1_genomic.fna.gz), using Burrows-Wheeler Aligner (BWA) software[25]. Mark Duplicates in Picard (<https://sourceforge.net/projects/picard/>) was used to eliminate the PCR duplication. We used Genome Analysis Toolkit (GATK) for base re-calibration and re-alignment near insertion or deletion regions[25]. SAMtools was used to estimate reference genome coverage of SNP and indel variations for identification and analysis.

RNA-seq analysis

Fruit at the full-red stage was collected from 'Linhuang No. 1' and 'Muzao' trees. Three biological replicates (five trees per replicates) of each sample were collected and stored at -80°C prior to RNA isolation. The total RNA from each sample was extracted from the fruit pericarp by the cetrimonium bromide (CTAB) method as described by Jaakola et al. (2001)[26]. The samples from the pericarp of each trees were mixed for RNA extraction. The effective concentration of the library was accurately quantified by quantitative real-time PCR (qPCR). Finally, cDNA libraries of the samples were sequenced on a flow cell using an Illumina HiSeq X-ten sequencing platform (BioMarker, Beijing, China). Transcriptome analysis was carried out by using reads mapping based on the reference genome. Gene expression levels were estimated using FPKM values (Fragments Per Kilobase of transcript per Million fragments mapped) and the false discovery rate (FDR) control method was used to identify the threshold of the P-value in multiple tests in order to compute the significance of any differences.

Gene cloning and protein sequence alignment

The target gene which encoding nitrite reductase (NiR) was amplified by cDNA from the pericarp of 'Linhuang No. 1' and 'Muzao' fruits. The primers used were forward primer ATGTCATCGTTCTCTGTTTCGGTTT and reverse primer TCAAAACGGGTGTTCCCTCGA. The amplified products were electrophoresed on 1% agarose gel and photographed with a gel imaging analysis system. The PCR products were recovered and purified according to the operation steps described in the Gel Recovery Kit (Takara, Japan). The recovered DNA fragment was ligated with a TA cloning vector and transformed into *Escherichia coli* DH5 α . After 12–16 h inverted culture at 37°C , single colonies were picked out and cultured in 2 ml lysogeny broth (LB) liquid medium. After 5–6 h shake culture (150 rpm) at 37°C , *Pst*I digestion was used for identification. The bacterial suspension was stored in a final concentration of about 25% (v/v) glycerin, and a 100- μl aliquot of the bacterial suspension was sent to Sangon Biotech (Shanghai, China) for sequencing. The cloned sequences were subjected to domain analysis by TBtools[27] and NCBI-CD Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Quantitative real-time PCR validation of nitrite reductase gene expression level

The quantitative real-time PCR (qPCR) was performed using the ChamQTM Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China). The cDNA templates were reverse transcribed using total RNA extracted from the pericarp of fruit at different development periods from 'Linhuang No. 1' and 'Muzao'. Then nitrite reductase gene (LOC107427052) was PCR amplified. The primers were: Forward: GCAAATCCGTGGTGTGGTT and Reverse: CAGCAAGAGGGTCCCAACT. The jujube Histon-H (HIS) gene (GenBank accession No. EU916201) was used as the internal reference for the gene expression analysis. The real-time PCR assay mix (20 μL) consisted of 2 μL cDNA sample (diluted 1:10), 10 μL 2 \times ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China), 0.6 μL of each primer (10 μM) and 6.8 μL distilled deionized H_2O . Each PCR assay was run on an iCycler iQ Real-Time PCR Detection System (Bio-Rad, CA, USA), using the method described in previous study[28]. The $2^{-\Delta\Delta\text{Cq}}$ method was used to calculate the relative amount of transcript template present in each PCR amplification mixture.

Determination of the nitrite reductase (NiR) activity and concentrations of nitrite and ammonia nitrogen

Determinations of the activity of nitrite reductase (NiR), and the concentrations of nitrite and ammonia nitrogen were carried out according to the Nitrite Reductase (NiR) Activity Assay Kit (BioBox, Beijing, China), Nitrite Content Assay Kit (BioBox, Beijing, China) and Plant Ammonia Nitrogen Content Assay Kit (BioBox, Beijing, China), respectively, following the manufacturer's instructions. The pericarp of fruits harvested from 'Linhuang No. 1' and 'Muzao' at different stages of fruit development was taken as samples. Independent analysis of each cultivar \times development stage combination was carried out three times.

Data analysis

SAS System 8.6 (SAS Institute, Cary, NC, USA) was used to analyze the data. Data analysis was carried out by one-way analysis of variance (ANOVA), with Student's t test being used to make multiple pairwise comparisons to determine significant differences among samples. Differences with $P < 0.05$ were considered to be statistically significant. Figures were constructed using Excel 2016, TBtools20 and Origin 9.0 (Microcal Software Inc., Northampton, MA, USA).

Declarations

Acknowledgements

We thank BioMarker (Beijing, China) and Lianchuan (Shanghai, China) for help with the transcriptome sequencing and genome resequencing respectively.

Authors' contributions

NL designed the experiment. NL, YS, JL, RH, XF and LL performed the experiments and analyzed the data. NL and LL wrote the manuscript. LL checked the manuscript. The author(s) read and approved the final manuscript.

Funding

This research was financially supported by the Shanxi Province Key Research and Development Projects (Grant number: 2020LYZY01) and Shanxi Agricultural University Science and technology innovation fund (Grant number: 2020BQ43 and 2020BQ28).

Availability of data and materials

All data generated or analyzed during this study are included in the main paper and supplementary information files. In addition, Whole-genome re-sequencing and RNA-seq data in this study were uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under accession number SRP307848 and SRP307646 respectively.

Ethics approval and consent to participate

The plant materials were obtained from Pomology Institute, Shanxi Agricultural University. Sampling of plant materials were performed in compliance with institutional, national, and international guidelines. We comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. The materials were publicly available for non-commercial purposes.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Horticulture, Shanxi Agricultural University, Taigu 030801, China. ²College of Forestry, Shanxi Agricultural University, Taigu 030801, China

Abbreviations

SNP: single nucleotide polymorphisms; indel: insertion/deletion variants; SV: structural gene variants; CNV: copy number variants; DEGs: differentially expressed genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; FPKM: Fragments per Kilobase Million; NiR: nitrite reductase; NR: nitrate reductase; CDS: coding sequence

References

1. Song J, Bi J, Chen Q, Wu X, Lyu Y, Meng X: **Assessment of sugar content, fatty acids, free amino acids, and volatile profiles in jujube fruits at different ripening stages.** *Food chemistry* 2019, **270**:344-352.
2. Peng S: **Current situation and Countermeasures of jujube production in China.** In: *The First National Symposium on Dry Fruit Production and Scientific Research Progress: 1998; Zanhuang, Heibei, China; 1998: 4.*
3. Li X, Tong J, Wang H, Sun W: **Ecological Adaptability of Ziziphus jujuba Mill. in North Shaanxi.** *Journal of Northwest Forestry University* 2000, **15**(02):13-18.
4. Guo Y, Shan G, Yang M: **Regionalization culture of jujube tree in China.** *China Fruits* 2002(04):46-48+53.
5. Lin J, Cheng Z, Xu M, Huang Z, Yang Z, Huang X, Zheng J, Lin T: **Genome re-sequencing and bioinformatics analysis of a nutraceutical rice.** *Molecular genetics genomics* 2015, **290**:955-967.

6. Aliko X, Maria M, Christos B, Christos K, Maria FA, Evangelos K, Michail M, Chrysanthi P, Antonios Z, Konstantinos K *et al.* **Whole genome re-sequencing of sweet cherry (*Prunus avium* L.) yields insights into genomic diversity of a fruit species.** *Horticulture Research* 2020, **7**(1).
7. Yu Y, Fu J, Xu Y, Zhang J, Ren F, Zhao H, Tian S, Guo W, Tu X, Zhao J *et al.* **Genome re-sequencing reveals the evolutionary history of peach fruit edibility.** *Nature communications* 2018, **9**(1).
8. Liu MJ, Zhao J, Cai QL, Liu GC, Wang JR, Zhao ZH, Liu P, Dai L, Yan G, Wang WJ: **The complex jujube genome provides insights into fruit tree biology.** *Nature Communications* 2014, **5**:5315.
9. Huang J, Zhang C, Zhao X, Fei Z, Wan K, Zhang Z, Pang X, Yin X, Bai Y, Sun X *et al.* **The Jujube Genome Provides Insights into Genome Evolution and the Domestication of Sweetness/Acidity Taste in Fruit Trees.** *PLOS Genetics* 2016, **12**(12).
10. Liu M, Wang J, Liu P, Zhao J, Zhao Z, Dai L, Li X, Liu Z: **Historical Achievements and Frontier Advances in the Production and Research of Chinese Jujube (*Ziziphus jujuba*) in China.** *Acta Horticulturae Sinica* 2015, **42**(09):1683-1698.
11. Guo M, Zhang Z, Li S, Lian Q, Fu P, He Y, Qiao J, Xu K, Liu L, Wu M *et al.* **Genomic analyses of diverse wild and cultivated accessions provide insights into the evolutionary history of jujube.** *Plant biotechnology journal* 2020.
12. Zhang C, Huang J, Yin X, Lian C, Li X: **Genetic diversity and population structure of sour jujube, *Ziziphus acidojujuba*.** *Tree Genetics & Genomes* 2015, **11**(1).
13. Guo M, Zhang Z, Cheng Y, Li S, Shao P, Yu Q, Wang J, Xu G, Zhang X, Liu J *et al.* **Comparative population genomics dissects the genetic basis of seven domestication traits in jujube.** *Horticulture research* 2020, **7**(1).
14. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD: **A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.** *Nature: International weekly journal of science* 2001, **409**(6822).
15. Zhaolei Z, Mark G: **Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes.** *Nucleic Acids Research* 2003, **31**(18):5338.
16. Wetterbom A, Sevov M, Cavelier L, Bergström TF: **Comparative Genomic Analysis of Human and Chimpanzee Indicates a Key Role for Indels in Primate Evolution.** *Journal of Molecular Evolution* 2006, **63**(5).
17. Edwards JD, Lee VM, McCouch SR: **Sources and predictors of resolvable indel polymorphism assessed using rice as a model** *Molecular Genetics and Genomics* 2004, **271**(3).
18. Britten RJ, Rowen L, Williams J, Cameron RA: **Majority of Divergence between Closely Related DNA Samples Is Due to Indels.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(8).
19. Long M, Betrán E, Thornton K, Wang W: **The origin of new genes: glimpses from the young and old.** *Nature Reviews Genetics* 2003, **4**(1).
20. Jin J, Yang L, Fan D, Liu X, Hao Q: **Comparative transcriptome analysis uncovers different heat stress responses in heat-resistant and heat-sensitive jujube cultivars.** *PLoS ONE* 2020, **15**(9):e0235763.
21. Zhang Q, Wang L, Liu Z, Zhao Z, Liu M: **Transcriptome and metabolome profiling unveil the mechanisms of *Ziziphus jujuba* Mill. peel coloration.** *Food chemistry* 2019, **312**:125903.
22. Zhou H, He Y, Zhu Y, Li M, Song S, Bo W, Li Y, Pang X: **Comparative transcriptome profiling reveals cold stress responsiveness in two contrasting Chinese jujube cultivars.** *BMC Plant Biology* 2020, **20**.
23. Silveira CM, Besson S, Moura I, Moura JGG, Almeida MG, Manessi-Zoupa E: **Measuring the Cytochrome c Nitrite Reductase Activity—Practical Considerations on the Enzyme Assays.** *Bioinorganic Chemistry & Applications* 2010, **2010**:1-8.
24. Ozawa K, Kawahigashi H: **Positional cloning of the nitrite reductase gene associated with good growth and regeneration ability of calli and establishment of a new selection system for *Agrobacterium* -mediated transformation in rice (*Oryza sativa* L.)** *Plant Science* 2005, **170**(2).
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.
26. Jaakola L, Pirttilä AM, Halonen M, Hohtola A: **Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit.** *Molecular Biotechnology* 2001, **19**(2):201-203.
27. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: **TBtools: an integrative toolkit developed for interactive analyses of big biological data.** *Molecular plant* 2020, **13**(8):1194-1202.
28. Li N, Fu L, Song Y, Li J, Li L: **Wax composition and concentration in jujube (*Ziziphus jujuba* Mill.) cultivars with differential resistance to fruit cracking.** *Journal of Plant Physiology* 2020, **255**:153294.

Tables

Table 1. Number of SNPs and indels, according to chromosome (Ch), detected for 'Linhuang No. 1' and 'Muzao', relative to the reference genome 'Dongzao' (A), and the comparison between them (B).

A) Chr	'Linhuang No.1'		'Muzao'		B) Linhuang No.1' vs 'Muzao'						
	SNP	InDel	SNP	InDel	SNP	InDel					
						≤10bp	11-20bp	21-40bp	41-80bp	81-160bp	≥161bp
NC_029679.1	429504	89039	430221	88650	47822	12968	1437	930	780	310	43
NC_029680.1	246814	54159	246954	53764	30426	8489	877	567	466	218	30
NC_029681.1	245843	54834	245635	54282	29879	8761	921	553	495	174	26
NC_029682.1	228631	48207	229082	47855	32212	8774	899	574	446	199	31
NC_029683.1	343538	69172	344093	68807	45243	12055	1303	776	615	222	40
NC_029684.1	217230	47474	217519	46967	25026	7109	812	510	409	167	31
NC_029685.1	251890	52302	252798	52,148	37682	9423	991	659	510	185	39
NC_029686.1	182760	42746	182578	42,337	22962	6699	688	445	389	169	29
NC_029687.1	240326	55181	239869	54,513	29621	8750	814	540	461	171	36
NC_029688.1	195933	41950	196006	41,515	22887	6562	733	471	358	149	28
NC_029689.1	189397	43820	189007	43,509	24530	6804	776	468	322	129	22
NC_029690.1	155137	35962	155452	35,687	18755	5599	630	393	287	137	29
unplaced genomic scaffold	696124	147012	697205	145,038	127817	34303	3223	1936	1323	510	60
Total	3623127	781858	3626419	775,072	494862	136296	14104	8822	6861	2740	444

Figures

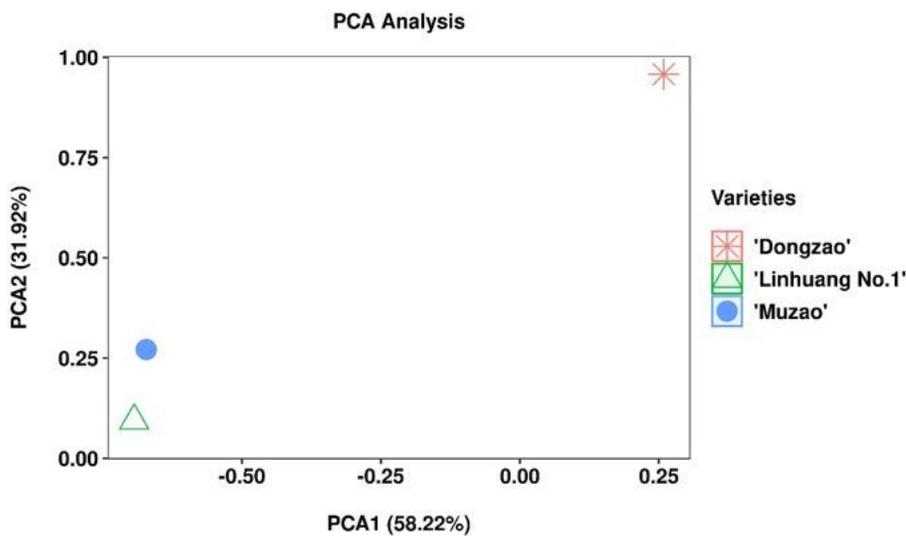


Figure 1
Principal component analysis (PCA) plot obtained from a similarity matrix among the *Ziziphus jujuba* Mill. accessions Linhuang No. 1 and Muzao, relative to the jujube genome reference cultivar Dongzao (ZizJuj_1.1, ID: 15586).

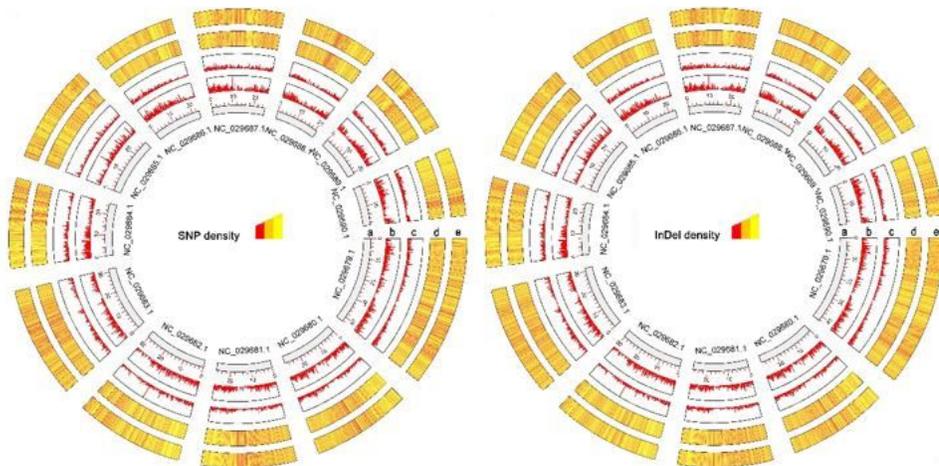


Figure 2

SNP and indel density plots by chromosome for 'Linhuang No. 1' and 'Muzao' against the reference genome 'Dongzao' (a: Chromosome scale; b: Gene density of chromosome positive chain; c: Gene density of chromosome negative chain; d: 'Linhuang No.1'; e: 'Muzao')

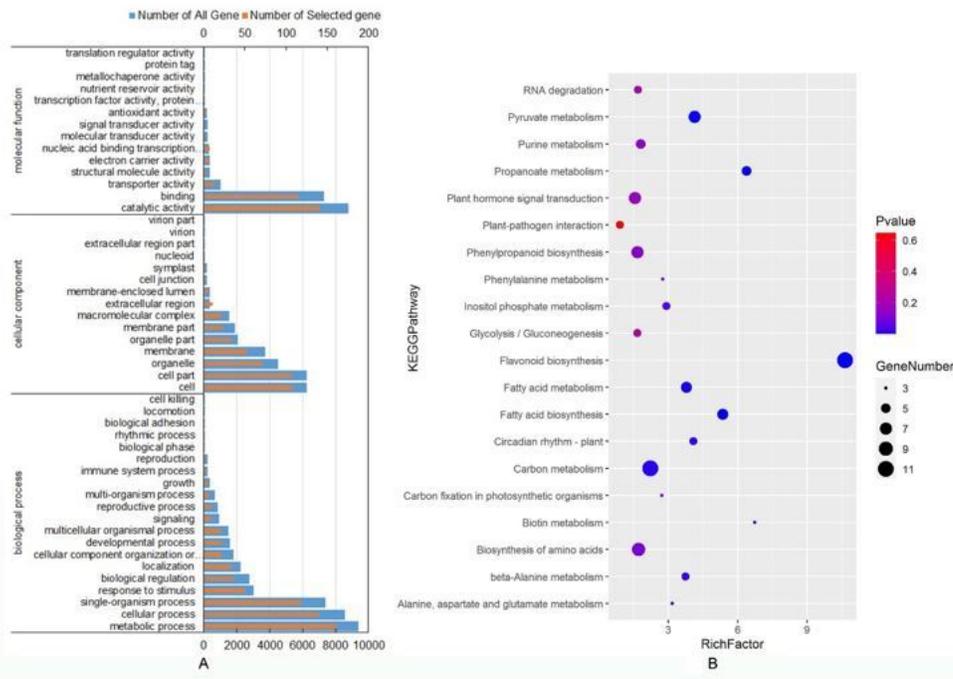


Figure 3

Gene Ontology (GO)(A) and Enrichment: Kyoto Encyclopedia (KEGG)(B) analysis of differentially expressed genes (DEGs)

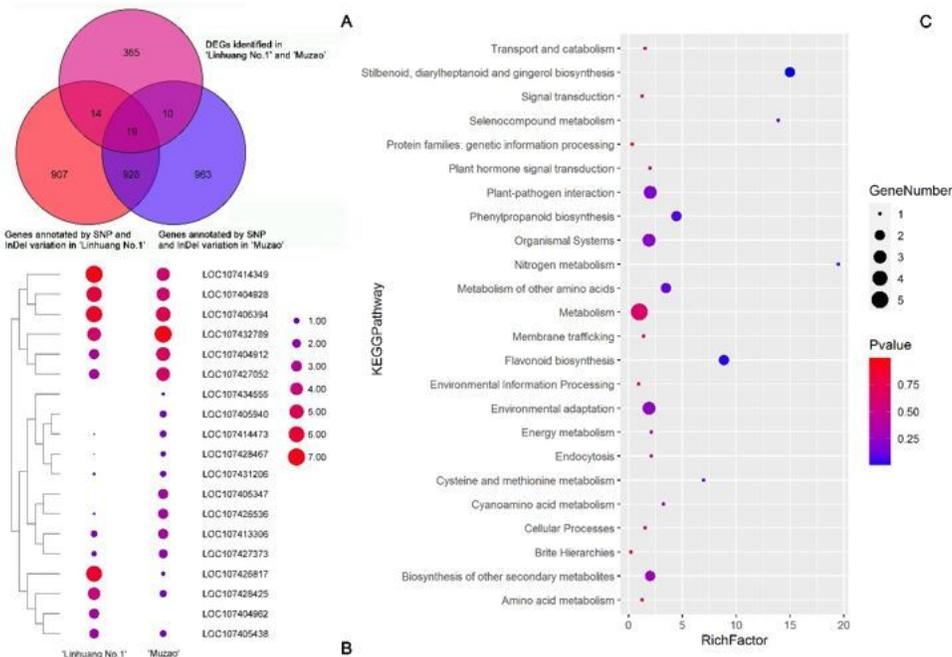


Figure 4

Venn diagram of 19 differentially expressed genes and SNP and indel annotated genes

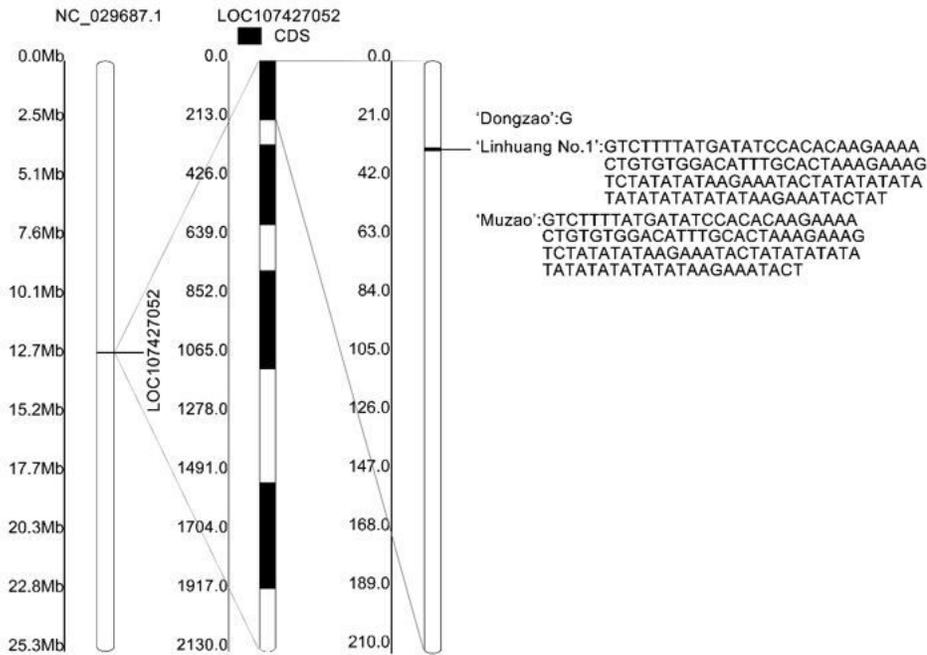


Figure 5

LOC107427052 mutation

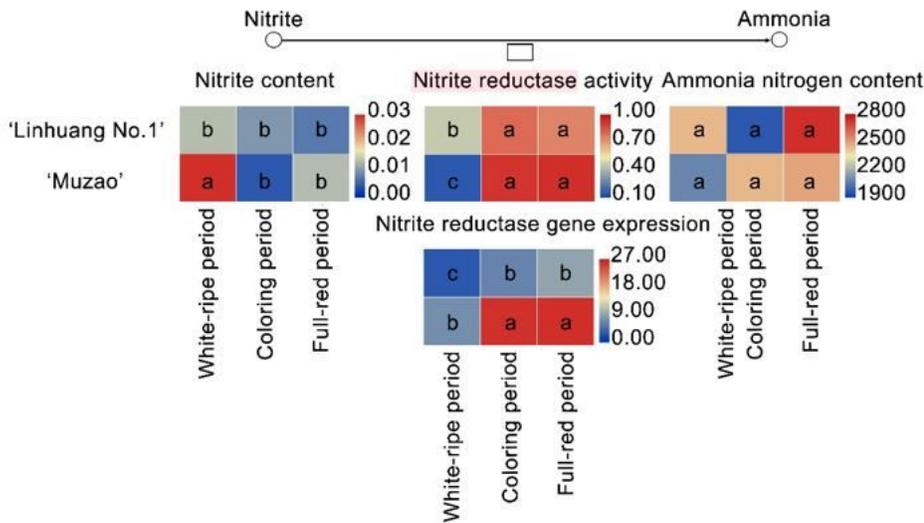


Figure 6

Levels of nitrite reductase (NiR) activity and gene expression, concentrations of nitrite and ammonia nitrogen at different stages of fruit development of the two cultivars. For a particular variable, any two samples with the same letter are not significantly different ($P > 0.05$)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Titlepageforthesupplementarymaterial.docx](#)
- [Additionalfile1TableS1.Summaryofsequencingandmappingresults.xlsx](#)
- [Additionalfile2FigureS1.StatisticsofsequencingdatafromjubeicultivarsLinhuangNo.1andMuzao.jpg](#)
- [Additionalfile3FigureS2.Vennmapofcultivarspecificityofexpressionofindividualgenes.jpg](#)
- [Additionalfile4TableS2.ThespecificexpressedgenesinLinhuangNo.1andMuzao.xlsx](#)
- [Additionalfile5FigureS3.AnalysisofdifferentiallyexpressedgenesDEGs.AupregulatedDEGsMuzaorelativetolinhuangNo.1BdownregulatedDEGsMuzaorel](#)
- [Additionalfile6FigureS4.SequencealignmentofDongzaoLinhuangNo.1andMuzaoLOC107427052gene.jpg](#)

- [Additionalfile7FigureS5.DomainpredictionofDongzaoLinhuangNo.1andMuzaoLOC107427052gene.jpg](#)