RESEARCH

Pangenome characterization and analysis of the *NAC* gene family reveals genes for *Sclerotinia sclerotiorum* resistance in sunflower (*Helianthus annuus*)

Yan Lu[1], Dongqi Liu[1], Xiangjiu Kong[1], Yang Song[1] and Lan Jing[1*]

Abstract

Background Sunflower (*Helianthus annuus*) is one of the most important economic crops in oilseed production worldwide. The different cultivars exhibit variability in their resistance genes. The NAC transcription factor (TF) family plays diverse roles in plant development and stress responses. With the completion of the *H. annuus* genome sequence, the entire complement of genes coding for NACs has been identified. However, the reference genome of a single individual cannot cover all the genetic information of the species.

Results Considering only a single reference genome to study gene families will miss many meaningful genes. A pangenome-wide survey and characterization of the NAC genes in sunflower species were conducted. In total, 139 *HaNAC* genes are identified, of which 114 are core and 25 are variable. Phylogenetic analysis of sunflower NAC proteins categorizes these proteins into 16 subgroups. 138 HaNACs are randomly distributed on 17 chromosomes. SNP-based haplotype analysis shows haplotype diversity of the *HaNAC* genes in wild accessions is richer than in landraces and modern cultivars. Ten *HaNAC* genes in the basal stalk rot (BSR) resistance quantitative trait loci (QTL) are found. A total of 26 *HaNAC* genes are differentially expressed in response to Sclerotinia head rot (SHR). A total of 137 *HaNAC* genes are annotated in Gene Ontology (GO) and are classified into 24 functional groups. GO functional enrichment analysis reveals that *HaNAC* genes are involved in various functions of the biological process.

Conclusions We identified *NAC* genes in *H. annuus* (HaNAC) on a pangenome-wide scale and analyzed *S. sclerotiorum* resistance-related NACs. This study provided a theoretical basis for further genomic improvement targeting resistance-related *NAC* genes in sunflowers.

Keywords Sunflower, *NAC* gene family, Pangenome characterization, Sclerotinia head rot resistance, Basal stalk rot resistance

*Correspondence:
Lan Jing
jinglan71@126.com
[1] College of Horticulture and Plant Protection, Inner Mongolia Agricultural University, Hohhot, China

Background

Many important cellular processes in plants are controlled by transcriptional regulation, such as signaling transduction, cellular morphogenesis, and various stress responses [1]. Regulation of gene expression requires a group of proteins known as transcription factors (TFs).

Transcription factors, which belong to a highly diverse family of proteins, generally function in protein complexes composed of multiple subunits. The *NAC* (NAM,

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 2 of 15

ATAF$_{1/2}$, and CUC$_2$) gene family encodes one of the largest and most important TFs in plants [2]. It has been reported that numerous NAC TFs take part in the regulation of a series of biological processes related to plant growth and development, including embryo and root development [3, 4], cell division [5], flowering [6], cell wall synthesis [7], leaf senescence [8], and response to abiotic and biotic stress [6, 9]. Because of their significance in plant complex life activities, genome-wide screening of NAC was performed in many plants like tobacco [10], poplar [11], cotton [12], foxtail millet [13], Arabidopsis, and rice [14].

As a large number of reference genomes have been released, genomic approaches can be employed to identify specific genes and study the correlations between candidate genes and heritable traits [15]. However, a single reference genome is unable to cover the full genetic information of a species due to structural variations which comprise deletions, insertions, translocations, inversions and duplications. These variations are often related to important agronomic traits [16, 17]. The reference genome of plant species is often derived from cultivated species, which cannot represent the rich genetic diversity of wild species, limiting the study of crop evolutionary and domestication history at higher breadth and depth. Therefore, conducting pangenomics analysis becomes crucial to ensure a comprehensive representation of genomic diversity within a species. Pangenomes have been created for many plant species, such as soybean [18, 19], maize [20], Brassica rapa [21], rice [22], Brassica oleracea [23], bread wheat [24], sunflower [25], and Brassica napus [26, 27].

The concept of a "pangenome" encompasses the entirety of genes in a species, without redundancy. It comprises two categories: core genes and variable genes. Core genes are found in all or nearly all individuals, while, variable genes occur only in certain individuals [16]. Variable genes include two variable types: copy number variations (CNVs) and gene presence/absence variations (PAVs) [28, 29].

Sunflower (Helianthus annuus L.) is an important source of edible oil and the seeds are used for food as well. It is produced worldwide because of its ability to grow and adapt in the most rigid environments. With the completion of the *H. annuus* genome sequence [30], the entire complement of genes coding for NACs has been identified and described [31]. The traditional reference genome (v1.0) provides a foundation for discovering these *NAC* genes. However, due to the effects of environmental factors, different individuals have formed extremely special genetic traits, and the reference genome of a single individual cannot represent the genetic diversity of the species in the process of evolution, leading to a loss of many meaningful genes.

In this study, *NAC* genes in *H. annuus* (HaNAC) on a pangenome-wide scale were identified, PAVs were detected, and the phylogenetic characteristics and distribution on chromosomes were analyzed. We studied single nucleotide polymorphisms (SNPs) and haplotype variation of *NAC* genes to understand the genetic diversity among different populations. In order to better understand the features of disease resistance-related NACs, the *NAC* genes in QTL regionfor Sclerotinia basal stalk rot (BSR) resistance were surveyed. Furthermore, we analyzed the expression of the *NAC* gene in response to Sclerotinia head rot (SHR). This study may provide clues in identifying disease resistance-related genes in this important crop.

Materials and methods

Pangenome

The *H. annuus* pangenome was described by Hübner et al. [25]. It was generated by sequencing 493 accessions, including 287 cultivated lines, 17 Native American landraces, and 189 wild accessions representing 11 compatible wild species.

Retrieval of *NAC* genes

Sunflower protein sequence data were obtained from the sunflower genome database (https://sunflowergenome.org/pangenome-data/). The Hidden Markov Model (HMM) profile of the NAC domain (PF02365) was downloaded from the Pfam database (http://pfam.xfam.org) and used to identify potential NAC genes from the sunflower genome via HMMER 3.3.2 software (http://www.hmmer.org/), with an E-value threshold of 10^{-5}. Only those proteins predicted to contain the NAM domain by the online tool SMART (http://smart.embl-heidelberg.de/smart/batch.pl) were considered as candidate NAC members. *Arabidopsis* NAC protein sequences were downloaded from the *Arabidopsis* genome, TAIR 10.0 release (http://www.arabidopsis.org/). Gene PAV was discovered across the cultivated gene pool using the SGSGeneLoss package [32].

Phylogenetic tree construction

The NAC protein sequences of sunflower and *Arabidopsis* were aligned using the program MAFFT 7.490 [33] with default parameters, and the tree construction was carried out by the Maximum Likelihood method using the program FastTree 2.1.11 (http://www.microbesonline.org/fasttree/). NAC family genes were classified into several subgroups based on similarities in NAC domain structures [34].

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 3 of 15

Density/distribution of the *NAC* gene on sunflower chromosomes

The NAC density/distribution was plotted using the rtracklayer package, karyoploteR package, and RColorBrewer package in Rscript (v4.0.3).

Gene–CDS–haplotype (gcHap) analysis of the *NAC* gene family

Vcftools 0.1.15 was used to screen the VCF files of Sunflower to obtain SNPs by further removing rare alleles with a missing rate of > 0.4 [35]. gatk_vcf_to_haplotype.pl (https://github.com/zhuochenbioinfo/VCF2HAP) was used to identify gcHapin all samples, and then to calculate the number of haplotypes in different populations.

To assess gene diversity across different populations, Shannon's equitability (E_H) [36] was calculated using the gcHap (gene–coding sequence–haplotype) data. The formula was:

$$E_H = \frac{1}{\ln N}(- \sum pi \ln pi)$$

where p_i is the proportion of the *i*th gcHap of a gene, N is the population size, and ln$N$ is the maximum possible diversity of a gene. E_H value ranges between 0 and 1.

Nei's genetic identity (I_{Nei}) [37] was used to measure the genetic differentiation among populations. For each gene, I_{Nei} between two populations was estimated with the gcHap data. The formula was:

$$I_{Nei} = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 Y_i^2}}$$

Xi and *Yi* represent the frequencies of the *i*th gcHap of a gene in populations X and Y, respectively.

E_H and I_{Nei} were visualized using the ggplot2 package and ggpubr package in R 4.0.3 [38].

Identification of NAC-genes conferring Sclerotinia BSR resistance

A total of six QTLs responsible for quantitative resistance to BSR have been identified in a sunflower recombinant inbred line population, one of each on linkage groups (LGs) 4, 9, 10, 11, 16, and 17 [39]. BLAST was used to compare the gene sequences at both ends of these QTLs with reference genome sequences (HA412-HO.v1.1). According to *HaNAC* and QTL positions in the reference genome, possible candidate *HaNAC* genes related to BSR are explored.

SNP data for the *H. annuus* genome and pangenome extra contigs were downloaded from (https://sunflowergenome.org/pangenome-data/HelianthusVariants.vcf.gz).

SNPs flanking the known Sclerotinia BSR resistance QTL regions were collected from the literature [39].

Waterfall plots were drawn using Variant Effect Predictor 88.13 [40], GenVisR 1.11.3 [41], vcftools 0.1.15 [42] and R 4.0.3.

Analysis of RNA-seq data of SHR

SHR is caused by the necrotrophic fungus *Sclerotinia sclerotiorum*. Fass et al. [43] studied gene expression at the early stages of infection (0, 4, and 8 dpi) in one susceptible (H89) and two tolerant inbred lines (HA853, RK416) inoculated with the pathogen in field conditions.

RNA-seq data were downloaded from NCBI and SRA accession number was SRP219154 [43]. The fastq-dump tool in the SRA Toolkit 2.10.0 (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=fastq-dump) was used to Convert SRA files to fastq files. Fastp 0.20.1 was used to trim low-quality bases (average Q-score below 20) and adaptor sequences in raw data [44]. The RNA-seq clean data of each sample were mapped to the sunflower pan-genome using HiSAT2 2.1.0 [45]. FPKM (Fragments Per Kilobase of exon model per Million mapped reads) value of *NAC* genes was calculated. DESeq2 1.32.0 [46] was used for differential expression analysis, |log2 fold change |≥ 1 and p < 0.05 were set as the threshold to determine differentially expressed genes. A heatmap of Log2 (FPKM + 1) values was generated using the ComplexHeatmap package (2.6.2, https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html) in R 4.0.3.

We analyzed the differential expression of *NAC* genes in each inbred line (IL)-time point combination. A total of 27 combinations was shown in Table S1.

Gene ontology (GO) annotation and enrichment analysis

All HaNAC sequences were compared against the sequences in the UniProt database using the BLASTP with an E-value cutoff of 1e-5. The Retrieve/ID Mapping tool was used (https://www.uniprot.org/uploadlists/) to convert UniProt IDs to GO IDs for HaNAC GO annotation. The agriGO was used for GO enrichment analysis (http://bioinfo.cau.edu.cn/agriGO/analysis.php). All genes in the sunflower pan-genome were used as background. GO enrichment results were visualized using Cytoscape 3.8.0 [47].

Results

Pangenome-wide identification of NAC family genes

A total of 139 NAC-encoding genes are identified in the *H. annuus* pangenome, including one *NAC* gene (*HaNAC139*) that is not present in the reference genome assembly. Of the 139 NAC genes, 114 (82.01%) are core genes (found across > 95% of the accessions) and 25 (17.99%) are dispensable genes of which 20 were found in > 5% ~ < 95% of the accessions, and 5 were rare genes

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 4 of 15

that were found in < 5% of the accessions. The detailed information on the 139 *NAC* gene sequences is shown in Table S2.

Phylogenetic analysis of NAC gene family in sunflower and *Arabidopsis*

To investigate the phylogenetic relationship among the HaNAC family members, a phylogenetic tree is constructed based on the alignment of 240 full-length protein sequences from sunflower and *Arabidopsis.* As indicated in Fig. 1, the 139 HaNACs are divided into 16

subgroups. Since AtNAC6, 23, 24, and 77 don't belong to any group, *HaNAC* genes highly similar to these genes are assigned to the Ha_NAC subgroup. The largest clade is the NAM subgroup containing 18 HaNACs, while the OsNAC8 subgroup constitutes the smallest clade with only one HaNAC98.

All subgroups of the HaNAC family contain variable genes (present only in some individuals) (Fig. 2). In total, the absence of 68 genes occurs in 5074 accessions (Table S3). In a total of 290 accessions, *HaNAC18* and *HaNAC55* are absent in 289 accessions, *HaNAC31* is
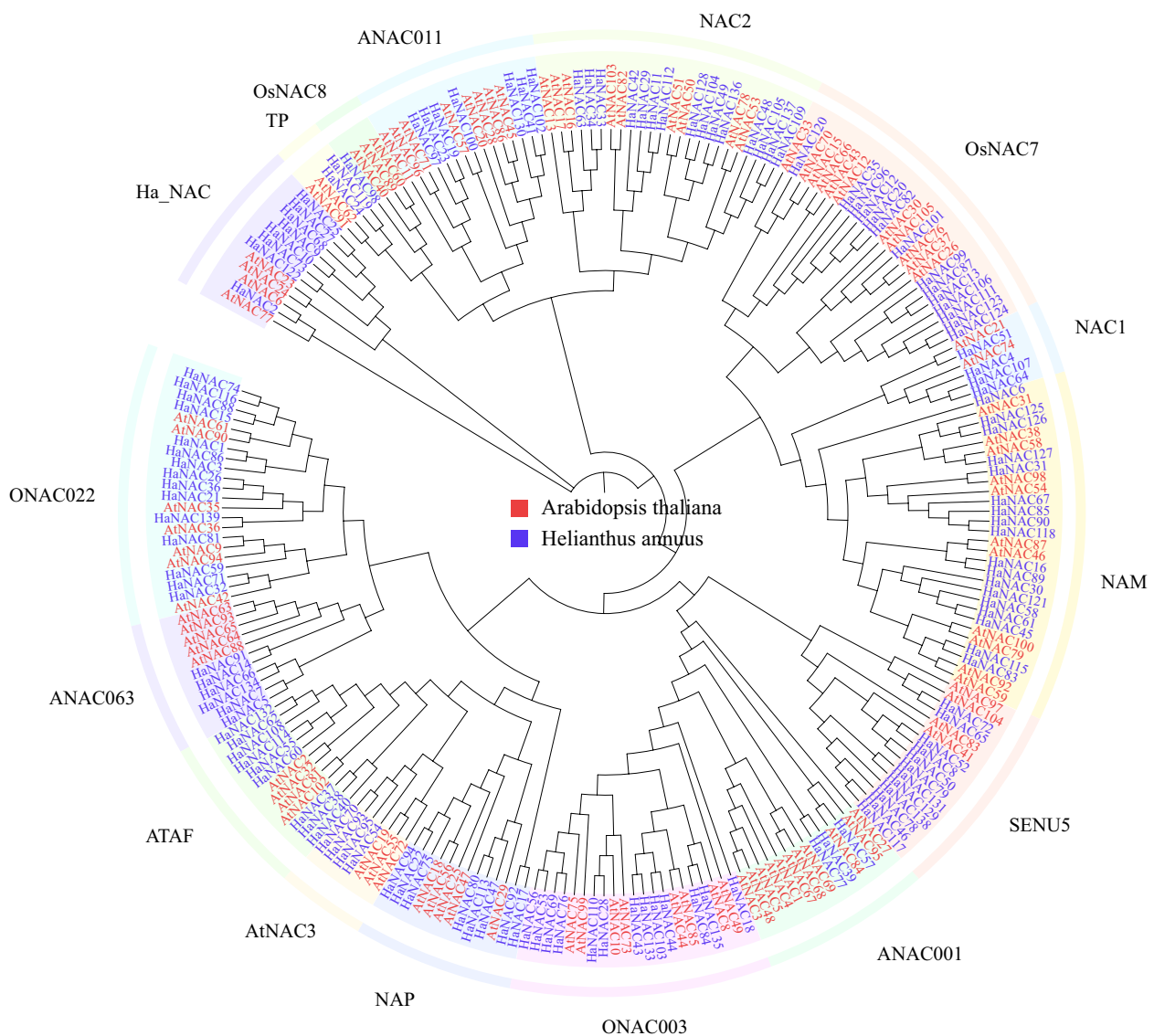
Fig. 1 A phylogenetic tree of sunflower and *Arabidopsis* NAC proteins. The amino acid sequences of NAC proteins were aligned using MAFFT 7.490, and a phylogenetic tree was generated using the Maximum Likelihood method of FastTree 2.1.11. NAC family genes were classified into different subgroups based on similarities in NAC domain structures [34]. Blue and red fonts denoted sunflower and *Arabidopsis* NACs, respectively. All NACs were classified into 16 subgroups (different colors for each clade)
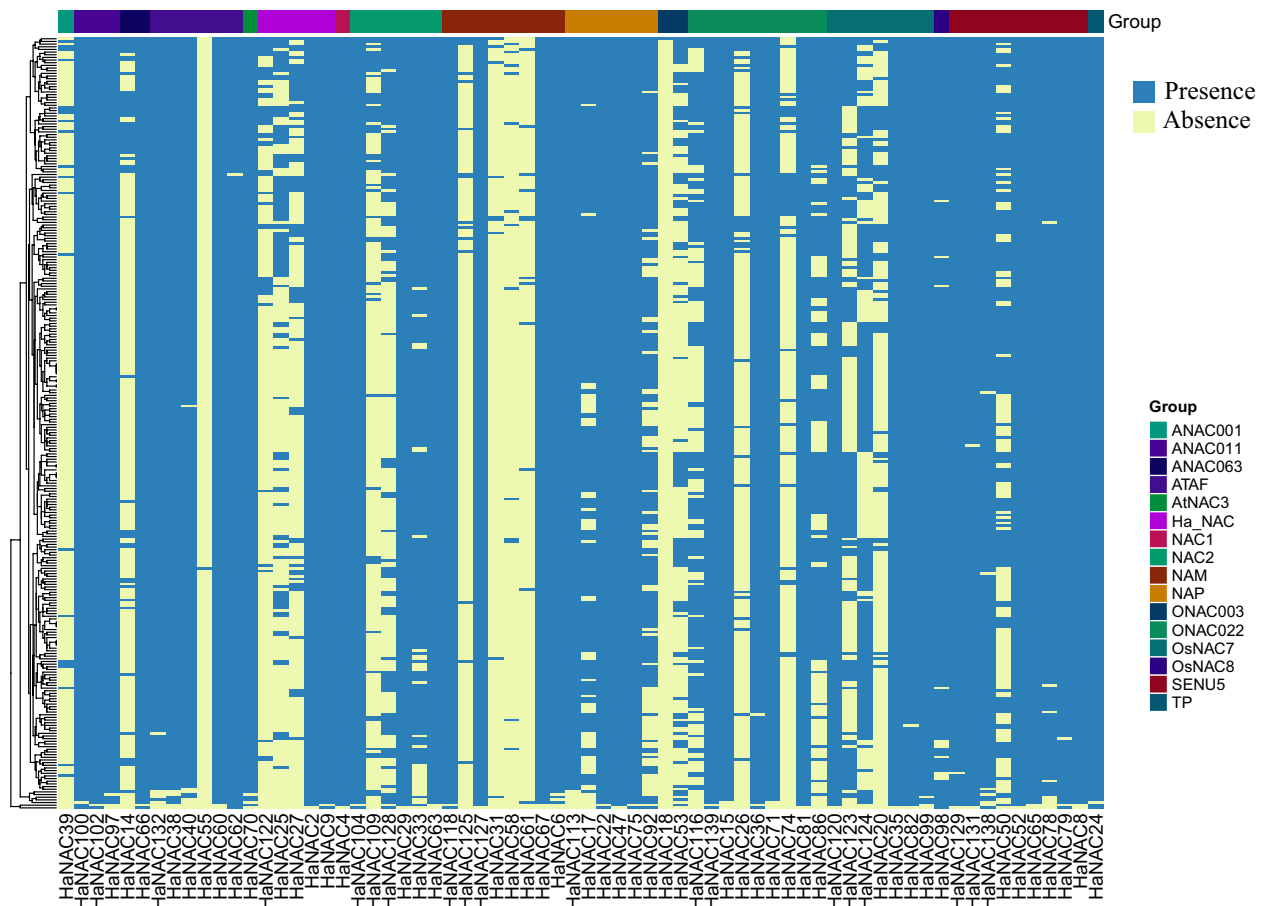
Fig. 2 Heat maps of gene presence/absence variation (PAV) in 290 sunflower lines

absent in 281 accessions, and *HaNAC61* is absent in 278 accessions.

The absence of variable genes in subgroups ANAC011, AtNAC3, NAC1, OsNAC8, and TP occur in few accessions, with 5, 3, 1, 11, and 1 respectively, indicating that these five subgroups are relatively stable among all groups. On the contrary, the gene absence in subgroups ANAC001 and ONAC003 occurs in a higher proportion of accessions, with 269 (~ 92.8%) and 484 (~ 83.4%), respectively, suggesting that these two groups are the most unstable among all groups and might endure strong selection pressure during sunflower domestication and breeding.
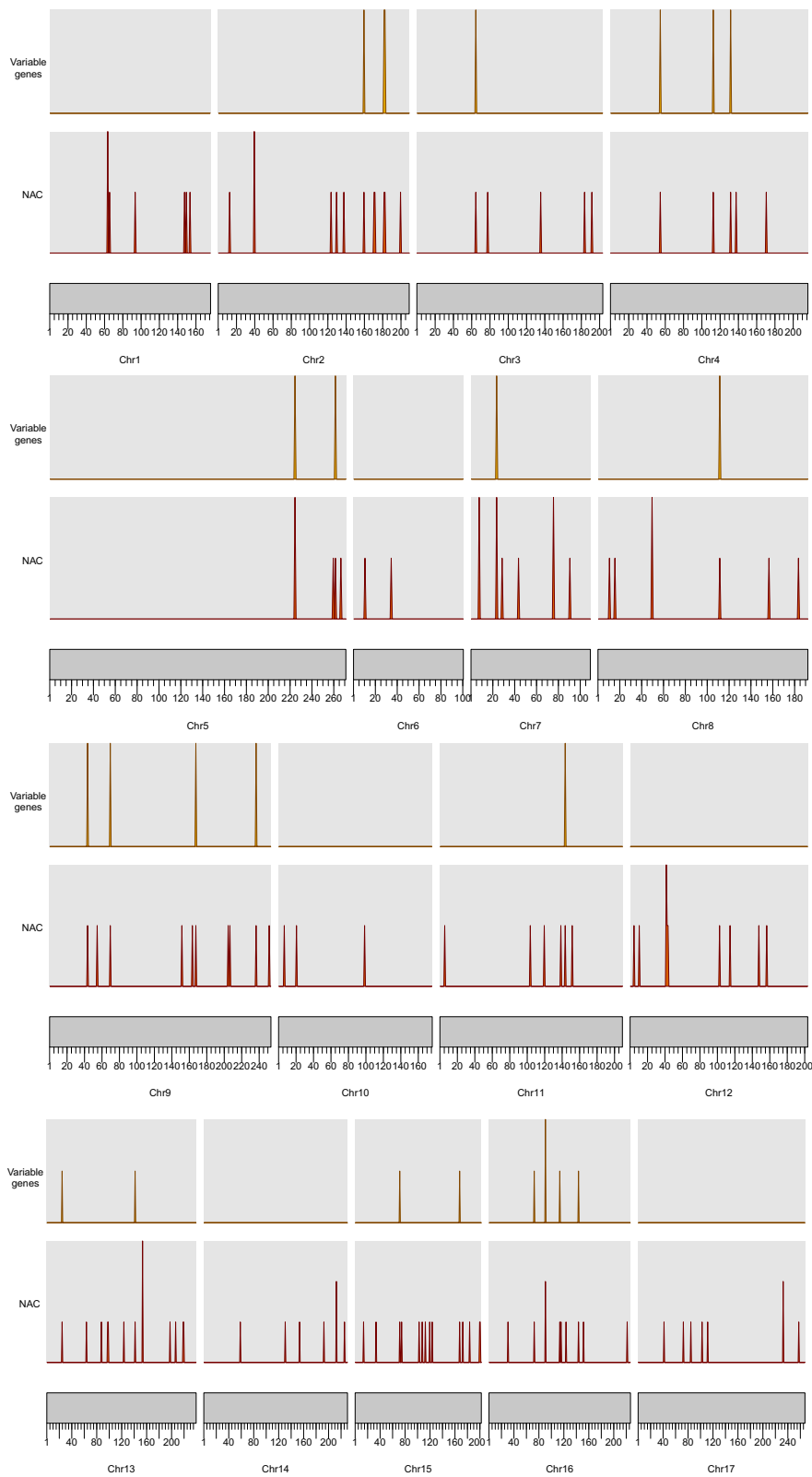
HaNAC gene distribution on sunflower chromosomes

We mapped the 138 *HaNAC* genes on all 17 chromosomes (Chr 1 to Chr17) and named them from *HaNAC1-HaNAC138* according to their chromosomal locations. As shown in Fig. 3, HaNAC sequences distribute unevenly over all chromosomes. Chr13 and Chr15 have a maximum of 14 *HaNACs* (~ 10.1%), respectively, whereas only 2 *HaNACs* (~ 1.4%) are located on Chr6. Chr5 has

the longest size of 271 Mb, but *NAC* genes are distributed only in the region between 220 and 271 Mb. Of the 138 *HaNACs*, 25 variable genes (dispensable genes and rare genes) are located on 11 chromosomes. No variable genes are located on Chr1, Ch6, Chr10, Chr12, Chr 14, and Chr 17, indicating that the *NAC* genes on these chromosomes are relatively stable during evolution and domestication. Half or more than half of the *NAC* genes on Chr4 and Chr16 are variable genes, suggesting that the *NAC* genes on these two chromosomes probably have undergone selection during sunflower domestication and diversification.

Analysis of gene–CDS–haplotypes in sunflower *NAC* gene family

We investigated the SNPs and haplotype constructed with adjacent SNPs in *HaNAC* genes. According to SNP information provided by Hübner et al. [25], 3247 SNPs are discovered within the CDS region of 108 *HaNACs*. No SNPs are found within the remaining 31 *HaNAC* genes (Table S4), which may be involved in some basic biological processes and serve as housekeeping genes.

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 6 of 15

Fig. 3 The distribution of *NAC* genes vs variable *NAC* genes on sunflower chromosomes. These densities were normalized by the genome-wide maximum of each measurement so that they peak at 1. The ruler at the bottom was used to show the size of each chromosome

Lu *et al. BMC Genomic Data* (2024) 25:39

Page 7 of 15

We made statistical analysis on the haplotypes in genes of different phylogenetic groups (Fig. 4A), and our results show that haplotypes are the most abundant in group ANAC011, while the least in group OsNAC8. The *NAC* gene in ANAC011 has the largest haplotype diversity, suggesting that these genes play different regulatory roles in different sunflower accessions. However, the *NAC* gene in OsNAC8 showed the least haplotype variation among different accessions, which may be due to its conserved function.

We calculated E_H to evaluate genetic diversity within populations (Fig. 4B, C). In general, genes with lower I_{Nei} values tend to make a greater contribution to population differentiation, while genes with higher I_{Nei} values have less impact. Furthermore, when the E_H value of a gene is lower, it suggests lower genetic diversity among individuals. Conversely, a higher E_H value indicates greater genetic diversity [35]. The E_H density distribution of HaNAC in wild accessions is closer to the right (larger E_H value) than that of other populations (Fig. 4B), indicating that haplotype diversity of the *HaNAC* gene in the wild accessions is richer. The peak value of E_H density in modern cultivars is closer to the left (lower E_H value), indicating that, the genetic diversity of the *HaNAC*

gene decreases after a long-term domestication. Figure 4C shows the E_H value of each population, and the mean value of the wild accessions is the largest, followed by landraces, whole and modern cultivars decreasing in their genetic diversity.

To compare the genetic diversity between different populations, we calculated I_{Nei} (Fig. 4D). Wild accessions vs landraces and wild accessions vs modern cultivars both have lower I_{Nei} values, indicating that the SNP-haplotypes of the *HaNAC* gene in the wild population are significantly different from those in the other two populations. However, the I_{Nei} values of landraces vs modern cultivars are mostly above 0.5, indicating that there is little haplotype difference between the *HaNAC* genes in these two populations.

Introgression in sunflower *NAC* gene family

According to the sunflower pangenome data [25], we verified whether the *NAC* gene family has introgression. The results show that HaNAC26 in cultivated sunflower association mapping (SAM) population has gene introgression from *H. argophyllus* and *H. neglectus*, and gene introgression from these two neighboring species into *H. annuus* arises in two samples.
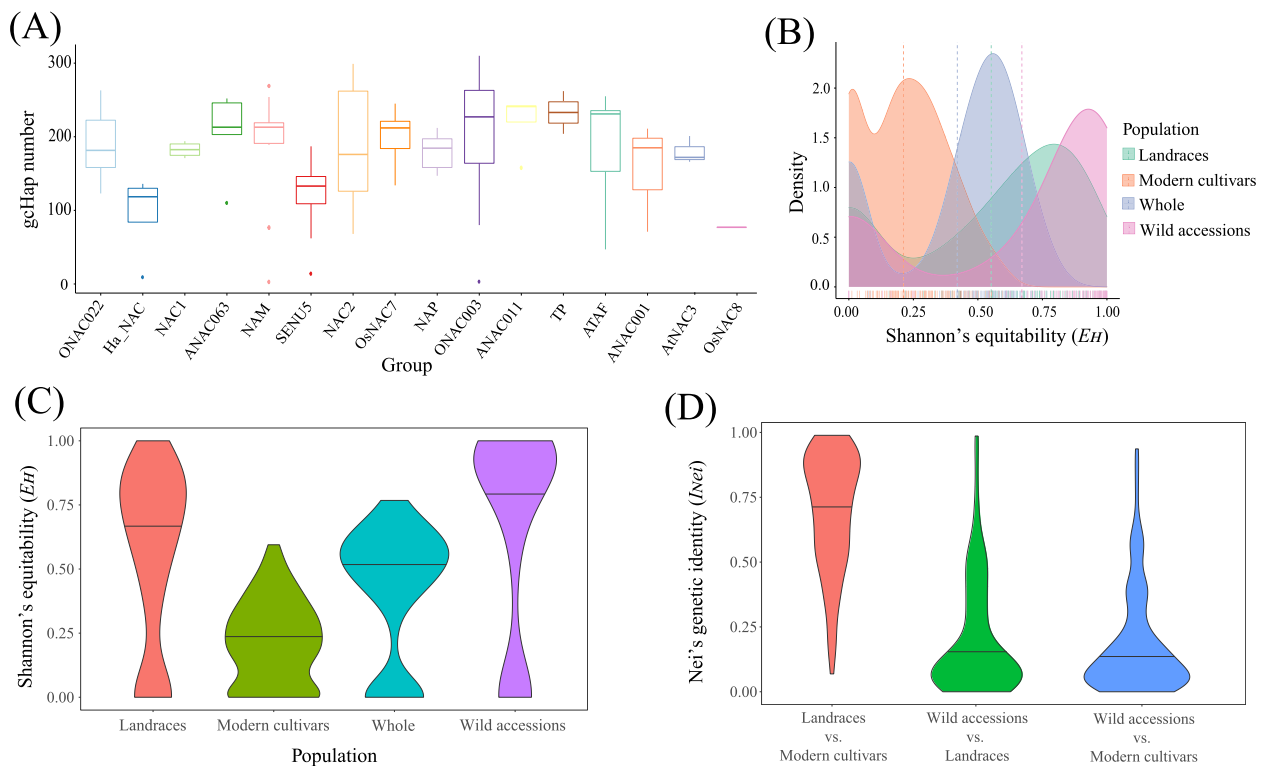
Fig. 4 gcHap numbers, Shannon's equitability (E_H) and Nei's genetic identity (I_{Nei}) of sunflower NAC family among different populations. **A** Distribution of gcHap number (gcHapN) of all 139 *NAC* genes in different phylogenetic groups. **B** Frequency distribution of E_H in landraces, wild accessions, modern cultivars and whole populations. **C** E_H distribution in four different populations. **D** I_{Nei} distribution of landraces vs modern cultivars, wild accessions vs modern cultivars, and wild accessions vs landraces

Lu *et al. BMC Genomic Data*　　　(2024) 25:39

Page 8 of 15

NAC genes in Sclerotinia BSR resistance QTL regions

Ten *HaNAC* candidates are identified at loci Qbsr-4.1, Qbsr-9.1, and Qbsr-16.1 (Table 1). *HaNAC26* locates at loci Qbsr-4.1, *HaNAC56* at loci Qbsr-9.1 and *HaNAC122*, *HaNAC123*, *HaNAC124*, *HaNAC125*, *HaNAC126*, *HaNAC127*, *HaNAC128*, *HaNAC129* at loci Qbsr-16.1. Of them, *HaNAC56*, *HaNAC126*, and *HaNAC127* are core genes, and the rest are dispensable ones. Combining with SNP and PAV information in the sunflower pangenome, the variation of 10 *HaNAC* genes in 492 accessions is analyzed (Fig. 5). The sunflower pan-genome contains the genetic information of 493 accessions. Because the *HaNAC* genes in the QTL regions have no SNP markers in accession PPN021, PPN021 is excluded from the analysis.

Table 1 The number of *HaNACs* at the Qbsr-4.1, Qbsr-9.1and Qbsr-16.1

Locus	Pseudomolecule	Start (bp)	End (bp)	Length (bp)	HaNAC number
Qbsr-4.1	Chr4	107675011	124808091	17133080	1 (HaNAC26)
Qbsr-9.1	Chr9	135225176	153762638	18537462	1 (HaNAC56)
Qbsr-10.1	Chr10	132441999	136480784	4038785	0
Qbsr-11.1	Chr11	170688742	185809916	15121174	0
Qbsr-16.1	Chr16	45544215	199280159	153735944	8 (HaNAC122、HaNAC123、HaNAC124、HaNAC125、HaNAC126、HaNAC127、HaNAC128、HaNAC129)
Qbsr-17.1	Chr17	150544245	230082427	79538182	0
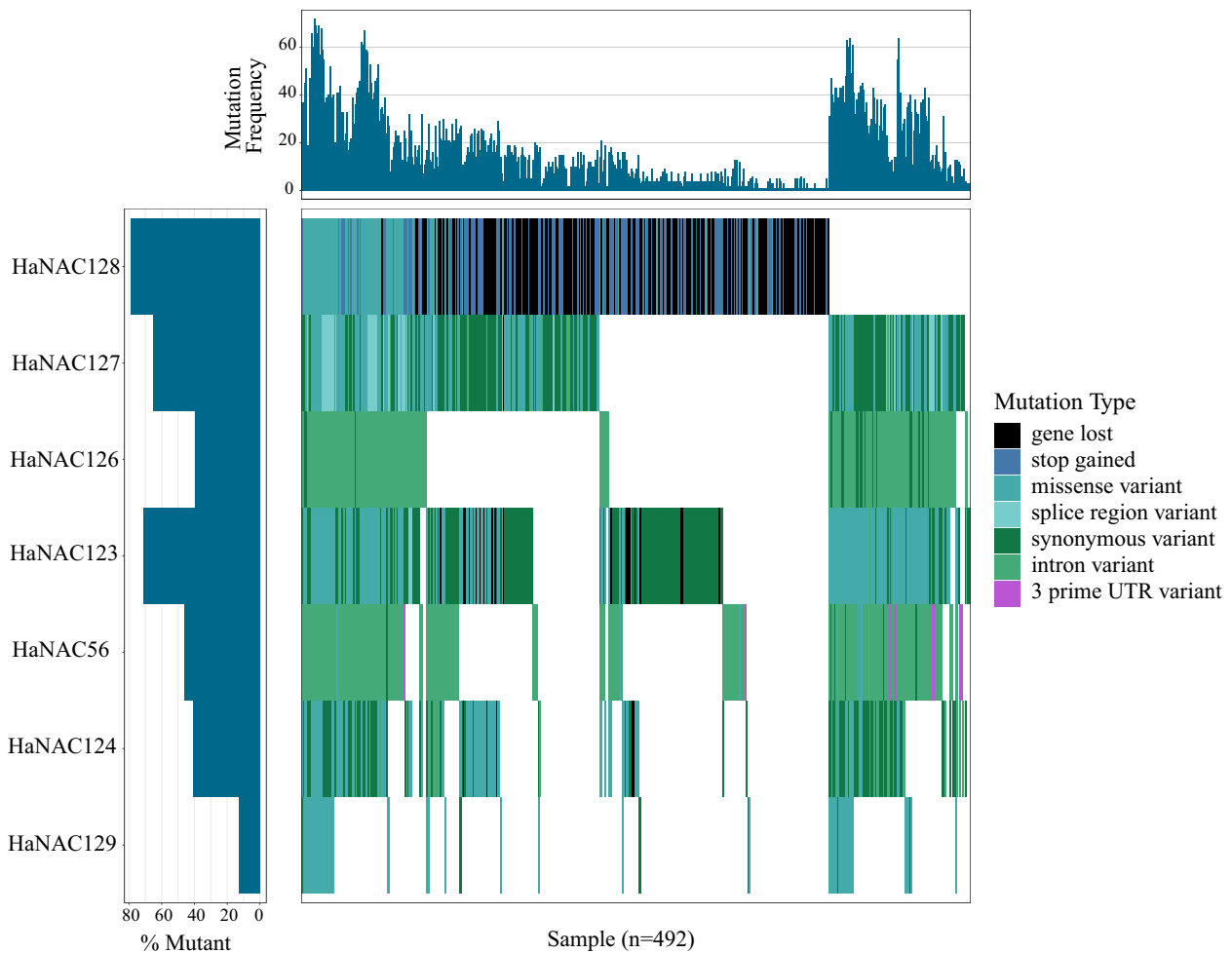
Fig. 5 Waterfall plot of *HaNAC* variation in the QTL region in 492 sunflower accessions

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 9 of 15

The genes *HaNAC26*, *HaNAC122*, and *HaNAC125* in the QTL region are not displayed in the waterfall map because they do not contain SNP markers, i.e., they do not have gene variation in 492 accessions, suggesting that these three genes are very conservative. Among the remaining seven *HaNAC* genes, *HaNAC128* shows the largest variant mainly due to the gene loss that is present in 171 samples, indicating that *HaNAC128* is subject to the greatest pressure of PAV selection. Mis-sense variant, synonymous variant, and intron variant are the main variation patterns of the *HaNAC* gene in the QTL region. *HaNAC123* has the highest proportion of synonymous variants, while *HaNAC126* has the highest proportion of intron variants. The *HaNAC129* shows low variation (<20%), but the majority of variants are missense, indicating that this gene has been strongly and positively selected in some accessions.

Expression analysis of *HaNAC* gene in response to SHR

Based on the transcriptome data of inoculated (I) and control (N) capitula of three sunflower inbred lines (ILs) [43], we analyzed the differential expression of each IL-time point combination.

A total of 26 *HaNAC* genes were expressed differentially (Fig. 6A, Table S5). *HaNAC* genes in groups

ANAC063 and NAC1 showed low expression levels. In addition, among all differentially expressed *HaNAC* genes, the number of $HaNAC_S$ in groups NAP and NAC2 was the largest, indicating that compared with others, the *HaNAC* genes in these two groups might be more involved in response to *S. sclerotiorum*.

All the combinations of I vs N had no differentially expressed gene (DEG) (data not shown). At three time points, HA853 and HA89 had larger numbers of DEGs in 8 dpi vs 0 dpi, 14 and 18 respectively, while 8 DEGs were found respectively in 4 dpi vs 0 dpi (Fig. 6B, C). However, RK416 had 10 DEGs in 8 dpi vs 4dpi, and no DEG in 4 dpi vs 0 dpi (Fig. 6D). The result indicated that HaNACs respond quickly to SHR in HA89 and HA853, while HaNACs respond slowly to SHR in RK416.

There were five DEGs in RK416_0_I vs HA89_0_I (Fig. 6E), seven in RK416_4_I vs HA89_4_I and nine in HA853_4_I vs RK416_4_I (Fig. 6F), five in RK416_8_I vs HA89_8_I and three in HA853_8_I vs RK416-8-I respectively (Fig. 6G), indicating that *HaNAC* genes in different lines responding to *S. sclerotiorum* are different at the same time point.

Most IL-time point combinations had larger numbers of up-regulated than down-regulated DEGs (Fig. 7A). Figure 7B shows the number of DEGs between three lines.
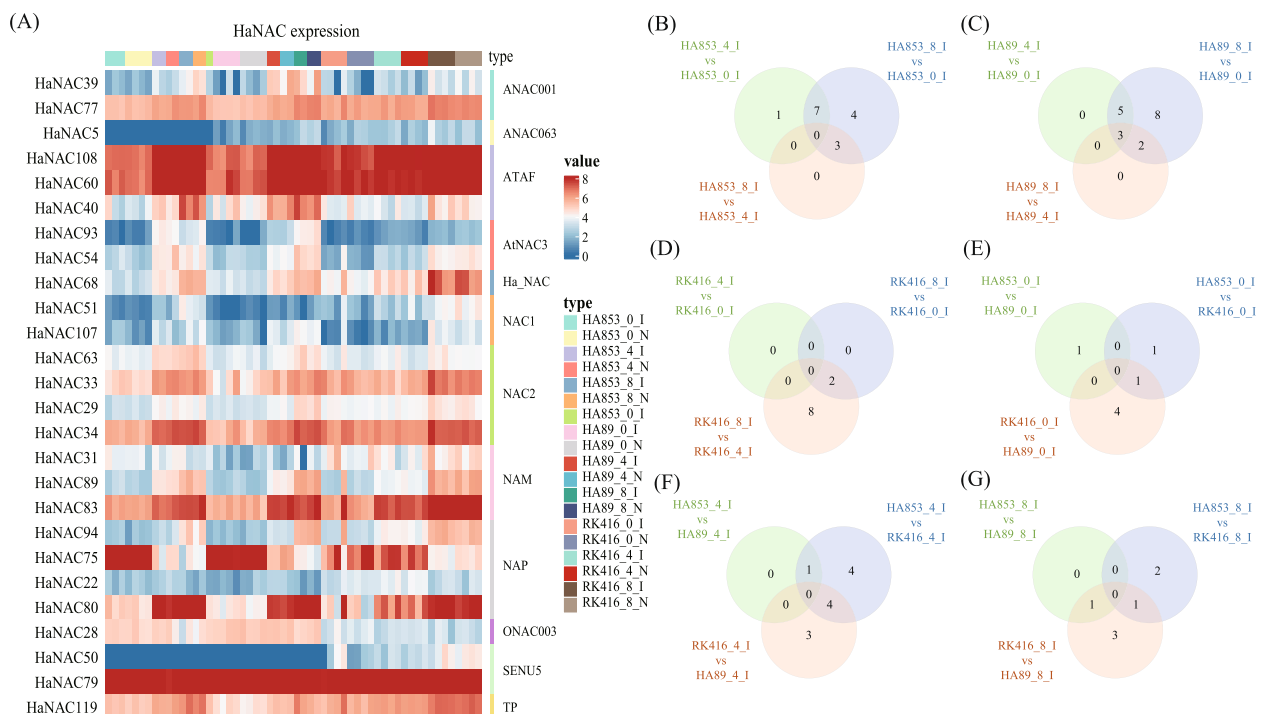
Fig. 6 Heatmaps and Venn diagrams. **A** Heatmaps of differentially expressed *HaNAC* genes. **B-G** Differentially expressed *HaNAC* genes in the three datasets through Venn diagrams software (available online: http://bioinformatics.psb.ugent.be/webtools/Venn/). Different colors meant different combinations
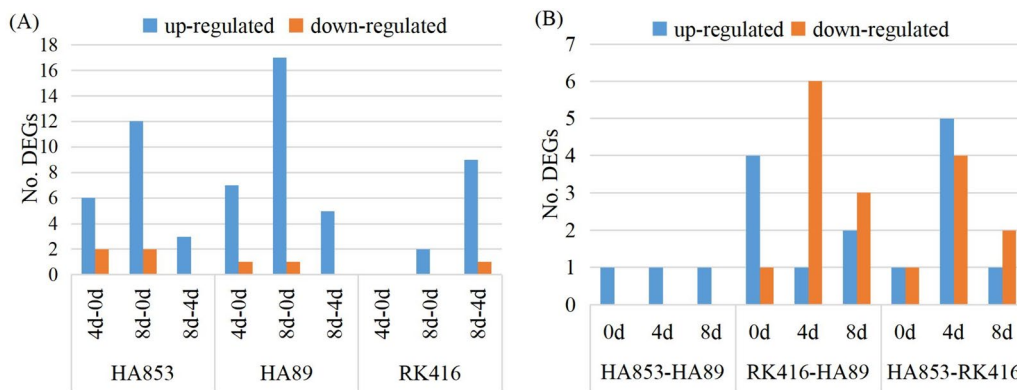
Fig. 7 Analysis of DEGs. **A** Number of up- and down-regulated DEGs between IL-time- inoculated samples. **B** Number of up- and down-regulated DEGs between three inoculated lines

Functional annotation and enrichment analysis of *HaNAC* genes

To further understand the function of *NAC* genes in sunflowers, we performed GO annotation and functional enrichment analysis for *HaNAC* genes.

A total of 137 HaNACs are annotated in Gene Ontology (GO) and are classified into 24 functional groups, including 15 groups in the biological process, six in cellular component, and three in molecular function (Fig. 8A). Within the biological process, the "metabolic process" (GO: 0008152) and "cellular process" (GO: 0009987) with 137 HaNACs respectively are predominant. In the category of cellular component, the three main groups are "organelle" (GO: 0043226, 137 HaNACs), "cell" (GO: 0005623, 137 HaNACs), and "cell part" (GO: 0044464,

137 HaNACs). The categories "binding" (GO: 0005488) and "transcription regulator activity" (GO: 0140110) are the most common in molecular function, represented by 137 and 114 HaNACs, respectively.

GO functional enrichment analysis reveals that *HaNAC* genes are mainly involved in the biological process (Fig. 8B). Figure 8C shows the network of enriched GO terms in the cellular component category. Figure 8D shows GO terms are enriched in the molecular function category, including DNA binding, transcription factor activity, sequence-specific DNA binding, nucleic acid binding transcription factor activity, nucleic acid binding, organic cyclic and heterocyclic compound binding. These represent the characteristics of the HaNAC proteins as transcription factors, which regulate gene expression via
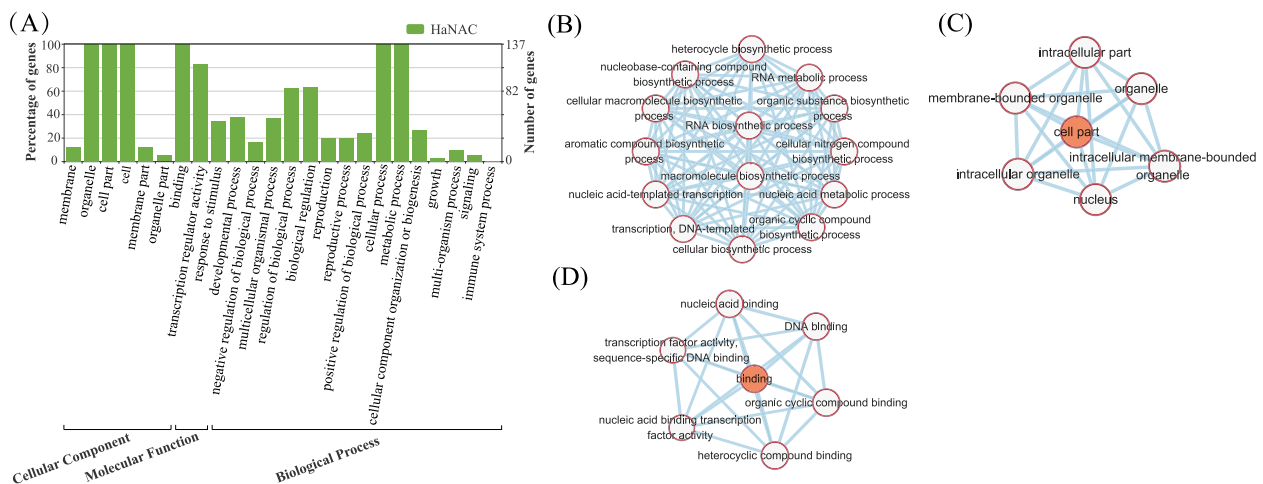
Fig. 8 Gene ontology annotation of HaNAC and network diagram of Go terms. **A** Gene ontology annotation of HaNAC. 137 sequences were grouped into three major functional categories and 24 sub-categories. **B** Network diagram of GO terms enriched in biological process. Only the top 10% of GO terms with the lowest FDR (false discovery rate) value were showed. **C** Network diagram of GO terms enriched in cellular component. **D** Network diagram of GO terms enriched in molecular function

transcription by binding to gene-specific sequences and affect the biological activities of cells.

Discussion

The NAC family is one of the largest plant-specific TFs that are involved in regulating growth, development, and stress responses. To date, *NAC* genes have been discovered in many plant species represented by 117 genes in *Arabidopsis* [14], 151 in rice [14], 101 in soybean [48], 145 in cotton [12], 104 in tomato [49], 148 in maize [50], 87 in sesame [51], 80 in watermelon [52], 145 in sorghum [53], 85 in sugarcane [54], 91 in cucumber [55], 104 in pepper [56], and 164 in cultivated peanut [57]. Li et al. [31] identified 150 HaNACs in sunflower through genome-wide survey (Ha412HO v1.0). The assembly of the cultivated sunflower pan-genome was guided by the HA412-HO. v1.1 reference sequence. Based on pan-genome data, we identified a total of 139 NACs with intact NAC domains in sunflower. Relatively numerous NACs in the sunflower were speculated to be highly involved in the complex transcriptional regulatory networks of sunflower. Multiple gene duplication events are believed to be responsible for this phenomenon as well.

According to the findings of Hübner et al. [25], the cultivated sunflower pan-genome consists of 61205 genes, with approximately 27% of these genes exhibiting variation across different genotypes. Our analysis found that 82% of 139 *NACs* in the *H. annuus* pangenome are core genes and 18% are dispensable genes (including rare genes). In plants, core genes often play a role in essential metabolic processes, while dispensable genes usually function in stress responses [19, 58–60] which tend to evolve faster under stronger selection [60, 61]. Previous studies have shown polymorphism level is higher in dispensable genes than in core genes [18, 26, 62, 63]. In the *B. oleracea* pangenome, nearly 20% of genes show PAV [23]. In the *Glycine soja* pangenome, dispensable genes account for 20% of the total genesets which exhibit greater variation than the core genome [18]. Sorghum pan-genome also displays large variation in genecontent, with 64% of gene families affected by PAV among genomes [64]. Through pangenomic analysis, we can study the retention and loss of genes during domestication and breeding [65]. In our study, PAV analyses revealed the loss of 5074 *NAC* genes during sunflower domestication and improvement, consistent with the trend found in sunflower domestication [25]. PAV is an important contributor to the studies of genetic diversity, gene identification, and molecular marker development in plants [66]. Understanding the PAV gene could support crop improvement applications and potentially reintroduce the gene into modern varieties [67].

Ooka et al. [34] classified NAC family proteins of *Oryza sativa* and *Arabidopsis thaliana* into two groups and 18 subgroups by sequence similarity. NAC proteins classified in the same groups may have similar functions in events common to monocotyledonous and dicotyledonous plants. Many findings suggest that the NAC proteins in subgroups NAM and NAC1 function in morphogenesis [2, 5], and proteins in the ATAF subgroup share a conserved role in stress responses [68]. NAC proteins in subgroup OsNAC3 (a monocot-specific subgroup) may be involved in monocot-specific responses to stress [34]. Analysis of the alignment of sunflower NAC proteins reveals that the proteins constitute a large family and belong to 16 subgroups (Fig. 1). Because proteins with domains similar in alignment are possible to have similar functions, our results will facilitate further functional analysis of sunflower NAC family genes. Our analysis shows the ATAF group consists of nine *NAC* genes in sunflower which may play a pivotal role in response to stress stimuli.

Sunflower holds global significance as it serves as a crucial oilseed crop, as well as a significant supplier of confectionery seeds and ornamental flowers. The cultivated sunflower are derived from wild *H. annuus* and were domesticated in what is now Central America earlier than 4000 years ago [69]. The wild sunflower is a potential source of cytoplasmic male sterility, and fertility restoration genes have been successfully introduced into cultivated sunflower [70–72]. Hübner et al. [25] reported approximately 10% of the cultivated sunflower pangenome contains introgression of the wild sunflower-derived gene, and 1.5% of the genes are introduced solely through introgression. Introgressed regions show an overrepresentation of genes associated with biotic resistance. Our analysis finds that introgression also exists in the sunflower *NAC* gene family. *HaNAC26* in the SAM population has gene introgression from two wild annual *Helianthus* species *H. argophyllus* and *H. neglectus*. *H. argophyllus* has been used as a valuable source of disease-resistance genes, which provide resistance against *Puccinia helianthi*, *Plasmopara halstedii*, and *S. sclerotiorum* in sunflower [73–75]. Hübner et al. [25] found that introgression is related to the formation of sunflower resistance ability. The introgression phenomenon of *HaNAC26* pointed out in this study reveals the *HaNAC26* is possibly associated with resistance.

The SNPs marker system is extensively used in modern genomics research [76]. Out of 139 *HaNAC* genes, 108 contained SNP markers. SNP-based haplotype analysis indicates that haplotype diversity of the *HaNAC* gene among wild accessions is richer than that in the landraces and modern cultivars implying

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 12 of 15

that selection pressure may lead to the loss of genetic diversity in certain populations during sunflower domestication. Low diversity may have weakened their ability to adapt to the environment. The sunflower gcHap diversity dataset generated in this study would contribute to sunflower basic research and future breeding. Polymorphisms within gene coding regions represent the most important part of the overall genetic diversity. Zhang et al. [35] characterized the gcHap diversity of 45963 rice genes in 3010 rice accessions. They found an average of 226 ± 390 gcHaps per gene in rice populations. Low frequencies of "favorable" gcHaps at most known genes related to rice yield in modern varieties suggest massive potential for improving rice by mining and pyramiding favorable gcHaps. The gcHap data were demonstrated to have greater power for detecting causal genes that affect complex traits. The rice gcHap diversity dataset would facilitate rice improvement in the future.

NAC transcription factors are known to be involved in coordinating responses to attacks by phytopathogens. Overexpression of the eggplant (*Solanum melongena*) transcription factor S*mNAC* suppresses resistance to bacterial wilt pathogen *Ralstonia solanacearum* [77]. Analysis suggests a putative NAC transcription factor Rph7 in barley (*Hordeum vulgare*) mediates the activation and strength of the basal defense response to leaf rust pathogen *Puccinia hordei* [78]. The rice *OsNAC30* mutant lines showed markedly reduced susceptibility to *Xanthomonas oryzae* pv. *oryzae* compared to wild-type plants. Mutation of *OsNAC59* conferred resistance to *Fusarium fujikuroi*, while mutation of *OsNAC101* increased susceptibility to this pathogen [79].

White mold caused by *S. sclerotiorum* is a devastating disease causing servere yield losses in sunflower production. Sunflower white mold has three different types of disease symptoms: BSR, mid-stalk rot (MSR), and head rot. So far, no major gene conferring complete resistance against this pathogen has been identified in cultivated sunflowers.

Crop breeding programs have faced challenges in identifying QTL that provide broad-spectrum resistance, which refers to resistance against various plant pathogens. These QTLs have proven to be elusive targets in breeding efforts. Six QTLs for resistance to BSR have been identified in the sunflower recombinant inbred line (RIL) population [39]. Our analysis data reveals that ten *HaNAC$_S$* are located at loci Qbsr-4.1, Qbsr-9.1, and Qbsr-16.1 probably play a regulatory role in BSR resistance. Three *NAC* genes (*HaNAC26*, *HaNAC122*, and *HaNAC125*) in the QTL region contain no SNP markers, therefore they have no variation in 492 accessions, which suggests a stable inheritance trait and valuable targets for breeders.

Identifying NAC candidates within QTL may help future breeding efforts in *H. annuus*. SNP markers tightly linked to resistance are also useful for breeding applications. Identifying both core and variable genes within these regions emphasizes the importance of employing pangenomics in these endeavors.

S. sclerotiorum has been reported to infect over 400 plant species [80]. Transcriptomic studies in *B. napus*, *A. thaliana,* and *Glycine max* have shown that defense against *S. sclerotiorum* involves transcription factor families, pathogenesis-related (PR) proteins, cell wall related proteins, as well as genes associated with cellular redox state, and hormone signaling pathways [81–85]. Joshi et al. [83] identified 30 TFs from *B. napus* post-infection with *S. sclerotiorum*, mainly including WRKY, NAC, ethylene response element binding factor (EREBF), MYBs, heat shock factors (HSFs), and C3H zinc finger. The results demonstrated the regulatory roles of plant TFs in response to pathogen challenges.

Fass et al. [43] investigated the transcriptional response of sunflowers to SHR. The analysis of differential gene expression revealed limited overlap among the ILs, indicating genotype-specific regulation of cell defense responses, potentially associated with variations in disease resistance strategies. All three ILs demonstrated an impact on the expression of genes related to cellular redox state and cell wall remodeling, aligning with existing understanding of the initiation of plant immune responses. Based on their data, we find a total of 26 differentially expressed *HaNAC* genes (~18.8% of the total *HaNAC* genes) involved in the defense against SHR. Our data analysis demonstrates the existence of diversified transcriptional responses to SHR within sunflower breeding lines and provides new evidence of the significant roles *HaNAC* genes played in response to pathogen challenges.

The GO functional annotation analysis further indicates that 'binding' and 'transcription regulator activity' are the most common molecular functions of HaNAC transcription factors, while 'cellular process' and 'metabolic process' are the most common biological processes, which is consistent with the characteristics of transcription factors.

Conclusion

In this study, we analyzed NACs in an *H. annuus* pangenome using a single reference and whole-genome sequencing data from 492 lines. Our various analyses reveal genomic landscape diversity and discover genes that have been lost during domestication in cultivated sunflowers. Our results highlight the potential of

variable genes to be used in genetic structural variation studies for future breeding programs. We identify some novel NACs that may contribute to resistance to Sclerotia white mold. Further genetic manipulation of these resistance-linked QTLs and genes will advance the precision breeding of sunflowers. Overall, the constructed sunflower pan-genome provides an important resource for sunflower improvement and gene discovery. And the findings will aid in furthering our understanding of not only the functions of core, and dispensable genes but also on various topics ranging from a better understanding of the evolutionary dynamics of gene families to genotype–phenotype associations.

Abbreviations

TF	Transcription factor
CNVs	Copy number variations
PAVs	Presence/absence variations
QTL	Quantitative trait loci
SHR	Sclerotinia head rot
GO	Gene Ontology
BSR	Basal stalk rot
HMM	Hidden Markov Model
gcHap	Gene–coding sequence–haplotype
E_H	Shannon's equitability
I_{Nei}	Nei's genetic identity
LGs	Linkage groups
FPKM	Fragments Per Kilobase of exon model per Million mapped reads
IL	Inbred line
SAM	Sunflower association mapping
DEG	Differentially expressed gene
MSR	Mid-stalk rot
RIL	Recombinant inbred line
PR	Pathogenesis-related
EREBF	Ethylene response element binding factor
HSFs	Heat shock factors

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-024-01227-9.

Additional file 1: Table S1. 27 Sample combinations for differentially expressed gene analysis.

Additional file 2: Table S2. List of NAC genes of sunflower retrieved from the sunflower pan-genome database.

Additional file 3: Table S3. The absence of 68 genes occurs in 5074 accessions.

Additional file 4: Table S4. The number of SNPs and haplotypes in sunflower NAC gene family.

Additional file 5: Table S5. 26. Differentially expressed *HaNAC* genes.

Authors' contributions

YL analyzed the data, drafted and edited the manuscript. DL analyzed and discussed the results for the *H. annuus* NAC family. XK and YS edited the manuscript and visualized the data. LJ conceived the study, edited and reviewed the manuscript, and connected with all authors, and involved them in major decisions about the publication. All authors have read and approved the content of the final manuscript.

Availability of data and materials

All data generated or analyzed in this study can be found in the supplementary information files and NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=4232.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000;290(5499):2105–10.
2. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The no apical meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell. 1996;85:159–70.
3. Larsson E, Sitbon F, Sundström J, von Arnold S. NAC regulation of embryo development in conifers. BMC Proc. 2011;5(Suppl 7):67.
4. Mao C, He J, Liu L, Deng Q, Yao X, Liu C, et al. OsNAC2 integrates auxin and cytokinin pathways to modulate rice root development. Plant Biotechnol J. 2020;18(2):429–42.
5. Kim YS, Kim SG, Park JE, Park HY, Lim MH, Chua NH, et al. A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*. Plant Cell. 2006;18(11):3132–44.
6. Zhang H, Cui X, Guo Y, Luo C, Zhang L. *Picea wilsonii* transcription factor NAC$_2$ enhanced plant tolerance to abiotic stress and participated in RFCP$_1$-regulated flowering time. Plant Mol Biol. 2018;98(6):471–93.
7. Zhong R, Richardson EA, Ye ZH. Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. Planta. 2007;225:1603–11.
8. Cao S, Zhang Z, Wang C, Li X, Guo C, Yang L, et al. Identification of a novel melon transcription factor CmNAC60 as a potential regulator of leaf senescence. Genes. 2019;10(8):584.
9. Zhang H, Kang H, Su C, Qi Y, Liu X, Pu J. Genome-wide identification and expression profile analysis of the NAC transcription factor family during abiotic and biotic stress in woodland strawberry. PLoS ONE. 2018;13:e0197892.
10. Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, et al. Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. Plant Physiol. 2008;147:280–95.
11. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol. 2010;10:145.
12. Shang H, Li W, Zou C, Yuan Y. Analyses of the NAC transcription factor gene family in *Gossypium raimondii* Ulbr.: chromosomal location, structure, phylogeny, and expression patterns. J Integr Plant Biol. 2013;55:663–76.
13. Puranik S, Sahu PP, Mandal SN, B VS, Parida SK, Prasad M. Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). PLoS One. 2013;8:e64594.

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 14 of 15

14. Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, et al. Genome-wide analysis of NAC transcription factor family in rice. Gene. 2010;465(1–2):30–44.

15. Qiu LJ, Xing LL, Guo Y, Wang J, Jackson SA, Chang RZ. A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol Biol. 2013;83:41–50.

16. Hurgobin B, Edwards D. SNP discovery using a pangenome: has the single reference approach become obsolete? Biology. 2017;6:21.

17. Zhao Q, Feng Q, Lu H, Li Y, Wang H, Tian Q, et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet. 2018;50:278.

18. Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol. 2014;32:1045–52.

19. Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, et al. Pan-Genome of wild and cultivated soybeans. Cell. 2020;182(1):162–76.

20. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, et al. Insights into the maize pangenome and pan-transcriptome. Plant Cell. 2014;26:121–35.

21. Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RG, et al. Beyond genomic variation–comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics. 2014;15:250.

22. Yao W, Li G, Zhao H, Wang G, Lian X, Xi W. Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biol. 2015;16:187.

23. Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, et al. The pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun. 2016;7:13390.

24. Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CKK, et al. The pangenome of hexaploid bread wheat. Plant J. 2017;90:1007–13.

25. Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nat Plants. 2019;5:54–62.

26. Hurgobin B, Golicz AA, Bayer PE, Chan CKK, Tirnaz S, Dolatabadian A, et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J. 2018;16:1265–74.

27. Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat Plants. 2020;6:34–45.

28. Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. Brief Funct Genomics. 2014;13:296–307.

29. Golicz AA, Batley J, Edwards D. Towards plant pangenomics. Plant Biotechnol J. 2016;14(4):1099–105.

30. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature. 2017;546:148–52.

31. Li W, Zeng Y, Yin F, Wei R, Mao X. Genome-wide identification and comprehensive analysis of the NAC transcription factor family in sunflower during salt and drought stress. Sci Rep. 2021;11:19865.

32. Golicz AA, Martinez PA, Zander M, Patel DA, Van De Wouw AP, Visendi P, et al. Gene loss in the fungal canola pathogen *Leptosphaeria maculans*. Funct Integr Genomics. 2015;15:189–96.

33. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

34. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakmi K, et al. Comprehensive analysis of NAC Family genes in *Oryza sativa* and *Arabidopsis thaliana*[J]. DNA Res. 2003;10(6):239–47.

35. Zhang F, Wang C, Li M, Cui Y, Shi Y, Wu Z, et al. The landscape of gene-CDS-haplotype diversity in rice (*Oryza sativa* L.): properties, population organization, footprints of domestication and breeding, and implications in genetic improvement. Mol Plant. 2021;14(5):787–804.

36. Sheldon AL. Equitability indices: dependence on the species count. Ecology. 1969;50:466–7.

37. Nei M. Genetic distance between populations. Am Nat. 1972;106:283–92.

38. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2016.

39. Talukder ZI, Seiler GJ, Song Q, Ma G, Qi L. SNP discovery and QTL mapping of sclerotinia basal stalk rot resistance in sunflower using genotyping-by-sequencing. Plant Genome. 2016;9(3):1–16.

40. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122.

41. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: genomic visualizations in R. Bioinformatics. 2016;32:3012–4.

42. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCF tools. Bioinformatics. 2011;27:2156–8.

43. Fass MI, Rivarola M, Ehrenbolger GF, Maringolo CA, Lia VV, et al. Exploring sunflower responses to *Sclerotinia* head rot at early stages of infection using RNA-Seq analysis. Sci Rep. 2020;10:13347.

44. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.

45. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.

46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

48. Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, et al. Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res. 2011;18(4):263–76.

49. Su HY, Zhang SZ, Yin YL, Zhu DZ, Han LY. Genome-wide analysis of NAM-ATAF1,2-CUC2 transcription factor family in *Solanum lycopersicum*. J Plant Biochem Biot. 2015;24:176–83.

50. Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, et al. Genomewide identification, classification and analysis of NAC type gene family in maize. J Genet. 2015;94:377–90.

51. Zhang Y, Li D, Wang Y, Zhou R, Wang L, Zhang Y, et al. Genome-wide identification and comprehensive analysis of the NAC transcription factor family in *Sesamum indicum*. PLoS One. 2018;13:e0199262.

52. Lv X, Lan S, Guy KM, Yang J, Zhang M, Hu Z. Global expressions landscape of NAC transcription factor family and their responses to abiotic stresses in *Citrullus lanatus*. Sci Rep. 2016;6:30574.

53. Kadier Y, Zu Y, Dai Q, Song G, Lin S, Sun Q, et al. Genome-wide identification, classification and expression analysis of NAC family of genes in sorghum [*Sorghum bicolor* (L.) Moench]. Plant Growth Regul. 2017;83:301–12.

54. Ramaswamy M, Narayanan J, Manickavachagam G, Athiappan S, Arun M, Gomathi R, et al. Genome wide analysis of NAC gene family 'sequences' in sugarcane and its comparative phylogenetic relationship with rice, sorghum, maize and Arabidopsis for prediction of stress associated NAC genes. Agri Gene. 2017;3:1–11.

55. Liu X, Wang T, Bartholomew E, Black K, Dong M, Zhang Y, et al. Comprehensive analysis of NAC transcription factors and their expression during fruit spine development in cucumber (*Cucumis sativus* L.). Hortic Res. 2018;5:31.

56. Diao W, Snyder JC, Wang S, Liu J, Pan B, Guo G, et al. Genome-wide analyses of the NAC transcription factor gene family in pepper (*Capsicum annuum* L.): chromosome location, phylogeny, structure, expression patterns, cis-elements in the promoter, and interaction network. Int J Mol Sci. 2018;19(4):1028.

57. Yuan C, Li C, Lu X, Zhao X, Yan C, Wang J, et al. Comprehensive genomic characterization of NAC transcription factor family and their response to salt and drought stress in peanut. BMC Plant Biol. 2020;20:454.

58. Danilevicz MF, Tay Fernandez CG, Marsh JI, Bayer PE, Edwards D. Plant pangenomics: approaches, applications and advancements. Curr Opin Plant Biol. 2020;54:18–25.

59. Jayakodi M, Schreiber M, Stein N, Mascher M. Building pangenome infrastructures for crop plants and their use in association genetics. DNA Res. 2021;28(1):1–9.

Lu *et al. BMC Genomic Data*   (2024) 25:39

Page 15 of 15

60. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell. 2021;184:3542–3558.e16.

61. Hübner S, Korol AB, Schmid KJ. RNA-Seq analysis identifies genes associated with differential reproductive success under drought-stress in accessions of wild barley *Hordeum spontaneum*. BMC Plant Biol. 2015;15:134.

62. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, et al. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. Nat Commun. 2017;8:2184.

63. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature. 2018;557:43–9.

64. Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, et al. Extensive variation within the pan-genome of cultivated and wild sorghum. Nat Plants. 2021;7(6):766–73.

65. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006;127(7):1309–21.

66. Bie H, Li Y, Zhao Y, Fang W, Chen C, Wang X, et al. Genome-wide presence/absence variation discovery and its application in Peach (*Prunus persica*). Plant Sci. 2023;335:111778.

67. Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D. Plant pan-genomes are the new reference. Nat Plants. 2020;6(8):1–7.

68. Christianson JA, Dennis ES, Llewellyn DJ, Wilson IW. ATAF NAC transcription factors: regulators of plant stress signaling. Plant Signal Behav. 2010;5(4):428–32.

69. Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH. Origin of extant domesticated sunflower in eastern North America. Nature. 2004;430:201–5.

70. Thompson TE, Zimmerman DC, Rogers CE. Wild *Helianthus* as a genetic resource. Field Crop Res. 1981;4:333–43.

71. Seiler GJ. Utilization of wild sunflower species for the improvement of cultivated sunflower. Field Crops Res. 1992;30:195–230.

72   Jan CC. Cytoplasmic male sterility in two wild *Helianthus annuus* L. accessions and their fertility restoration. Crop Sci. 2000;40:1535–8.

73. Seiler GJ. Utilization of wild *Helianthus* species in breeding for disease resistance. Proceedings of the International Sunflower Association (ISA) Symposium "Sunflower breeding on resistance to diseases," June, 23-24, 2010, Krasnodar, Russia. p. 36–50.

74. Qi LL, Foley ME, Cai XW, Gulya TJ. Genetics and mapping of a novel downy mildew resistance gene, Pl18, introgressed from wild Helianthus argophyllus into cultivated sunflower (*Helianthus annuus* L.). Theor Appl Genet. 2016;129:741–52.

75. Qi L, Long Y, Talukder ZI, Seiler GJ, Block CC, Gulya TJ. Genotyping-by-sequencing uncovers the introgression alien segments associated with sclerotinia basal stalk rot resistance from wild species—I. *Helianthus argophyllus* and *H. petiolaris*. Front Genet. 2016;7:219.

76. Ganal MW, Altmann T, Röder MS. SNP identification in crop plants. Curr Opin Plant Biol. 2009;12:211–7.

77. Chen N, Wu S, Fu J, Cao B, Lei J, Chen C, et al. Overexpression of the eggplant (*Solanum melongena*) NAC family transcription factor S*mNAC* suppresses resistance to bacterial wilt. Sci Rep. 2016;6:31568.

78. Chen C, Jost M, Outram MA, Friendship D, Chen J, Wang A, et al. A pathogen-induced putative NAC transcription factor mediates leaf rust resistance in barley. Nat Commun. 2023;14:5468.

79. Son S, Song G, Nam S, Lee G, Im J, Lee KS, et al. CRISPR/Cas9-mediated mutagenesis of rice *NAC* transcription factor genes results in altered innate immunity. Plant Physiol. 2024;00:1–5.

80. Boland GJ, Hall R. Index of plant hosts of *Sclerotinia sclerotiorum*. Can J Plant Pathol. 1994;16:93–108.

81   Dai FM, Xu T, Wolf GA, He ZH. Physiological and molecular features of the pathosystem *Arabidopsis thaliana* L.-*Sclerotinia sclerotiorum* Libert. J Integr Plant Biol. 2006;48:44–52.

82. Zhao J, Wang J, An L, Doerge RW, Chen ZJ, Grau CR, et al. Analysis of gene expression profiles in response to *Sclerotinia sclerotiorum* in *Brassica napus*. Planta. 2007;227:13–24.

83   Joshi RK, Megha S, Rahman MH, Basu U, Kav NNV. A global study of transcriptome dynamics in canola (Brassica napus L.) responsive to *Sclerotinia sclerotiorum* infection using RNA-Seq. Gene. 2016;590:57–67.

84. Girard IJ, Tong CB, Becker MG, Mao X, Huang J, de Kievit T, et al. RNA sequencing of *Brassica napus* reveals cellular redox control of *Sclerotinia* infection. J Exp Bot. 2017;68:5079–91.

85. Wen Z, Tan R, Zhang S, Collins PJ, Yuan J, Du W, et al. Integrating GWAS and gene expression data for functional characterization of resistance to white mould in soya bean. Plant Biotechnol J. 2018;16:1825–35.

Publisher's Note