

# Comparative transcriptomes and genome-wide identification reveal salt stress-responsive PP2C in Jute (*Corchorus capsularis*)

**Aminu Kurawa Ibrahim**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Yi Xu**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Qingyao He**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Sylvain Niyitanga**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Muhammad Zohaib Afzal**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Lilan Zhang**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Jianmin Qi**

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

**Liwu Zhang** (✉ [zhang\\_liwu@hotmail.com](mailto:zhang_liwu@hotmail.com))

Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

---

## Research Article

**Keywords:** Transcriptome sequencing, protein phosphate 2C (PP2C), abscisic acid, transcriptional factors, plant hormone signal transduction, salt stress tolerance

**DOI:** <https://doi.org/10.21203/rs.3.rs-190737/v1>

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

---

# Abstract

**Background:** The jute plant is of great significance and economic relevance to humanity, but its production has been hindered due to abiotic influences, especially salt stress. Hitherto, the molecular bases for this vital feature await future exploration. The *abscisic acid* (ABA) signaling pathway comprises many regulated genes and plays a role in plant response to stress, however, a balance between the multiple pathways is always needed for any plant developmental process. In this study, we used a transcriptomic approach to unveil the molecular bases behind this trait. Salt tolerant (J194) and sensitive (J7) germplasms were subjected to sodium chloride (NaCl) stress at a different time point, from which leaf and roots samples were taken for transcriptome analyses.

**Result:** The plant hormone signal transduction pathway was the most abundant observed in the study; the Pyrrolysine (PYL) gene (Cc.03G0016680) was up-regulated, which supports the basic model of abscisic acid (ABA). The quantitative reverse transcription-PCR (qRT –PCR) and the correlation analysis validated the Ribonucleic acid sequence (RNA-seq) results. The candidate genes' relative expression level was higher in J194, especially in protein phosphate 2C (PP2C). *Corchorus capsularis* PP2C gene family revealed 38 members, phylogenetic analysis categorized PP2C into 15 based upon conserved domains. Eleven conserved motifs were identified, and most of the genes had the same number of conserved motifs. The exon-intron ranges of (3-21) and (2-20), respectively. Moreover, among the plant hormone signal transduction pathway PP2C genes, Cc.03G0016550 and Cc.07G0028160 were up-regulated in J194 root tissues at 6-hour exposure NaCl, as such recommended to be salt-tolerant candidature genes. It was noted that most of the *Corchorus capsularis* PP2C genes were involved from segmental duplication, and analysis of the key stress marker salt-tolerant PP2C genes validated the salt tolerance individuals.

**Conclusion:** These results provided valuable insight into salt tolerance transcriptome and indicated that PP2C had provided a stepping-stone to the molecular mechanism in *Corchorus capsularis*. Furthermore, differentially expressed genes, motifs, gene structure, and the chromosomal location of salt tolerance candidate genes might have experienced functional divergence. As such, their further study will enhance salt tolerance in *Corchorus capsularis*.

## Introduction

*Corchorus olitorius* L and *Corchorus capsularis* L are the essential fiber crops cultivated in the world [1]. The demand for these crops increases globally due to their eco-friendly and broad-spectrum application characteristics [2]. However, their production has been affected by salt stress, which is a major environmental factor that hinders plant growth and development and affects overall yield [3]. It can exert both primary (shift in ion contents or osmotic dynamics, leading to ion toxicity) and secondary (includes oxidative stress, damage to cellular components, and metabolic dysfunction) effects on the plant. As such, the affected plants activate a flow of signal transduction pathways that alter gene expression, membrane trafficking, hormonal levels, and protein phosphorylation [4]. A wide range of physiological, biochemical, and molecular processes in plants are affected by salt stress [5]. Salt stress also regulates the levels of Reactive Oxygen Species (ROS) as well as changes in the expressions of genes, particularly those of the

Transcription Factors (TFs) [6]. When salt-tolerant and sensitive genotypes are compared, plant hormonal signal transduction and arginine and proline metabolisms are the most significant pathways to be considered. The phytohormone abscisic acid (ABA) is among the salt-stress-induced compound central to salt stress responses [3]. Firstly, the plasma membrane nicotinamide adenine dinucleotide phosphate oxidase (NADPH oxidase) generates Hydrogen peroxide ( $H_2O_2$ ) to modulate calcium signals that affect the downstream ABA responses [7]. ABA binds to Pyrabactin resistance 1-like (PYR/PYL) receptors and 2C protein phosphatase (PP2C proteins), forming a PYL-ABA-PP2C complex that activates Serine/threonine-protein kinase SRK2 (SnRK2s) [8]. Additionally, antioxidant production is regulated through a group of protein kinases phosphorylates bZIP and other TFs [9]. Moreover, mitogen-activated protein kinase (MAPK) pathways are activated through ABA, directly phosphorylating many other ABA effector proteins [7].

A balance between the multiple pathways is always needed for any plant developmental process, which can easily be compromised under abiotic stress influences [10]. Production and balance of ROS are related to photosynthesis, phenylalanine metabolism, and peroxidase pathways [11]. The photosynthetic pathway's electrons leak to  $O_2$  underlying stress, resulting in a massive generation of ROS. However, numerous effector genes are involved in peroxidase pathways such as superoxide dismutase (SOD), ferulic acid [12], catalase (CAT) [10], and isocitrate dehydrogenase (ICDH) present differential expression. If the pathways can preserve ROS at a relatively low level and reconstruct the new balance, the tissues and cells may be prevented from damage or even death due to oxidation. Peng et al. [13] investigated salt-tolerant and sensitive cotton genotypes by transcriptomic analysis and reported up-regulated genes related to abscisic acid, transporter, ethylene, as well as membrane receptors signal transduction [13]. Improving salt tolerance in jute will enhance its cultivation in saline environments. However, many previous studies concentrated on morphological, physiological, and proteomic analyses, ignoring transcriptomic approaches [7].

PP2C modulate stress-signaling pathways and reverse the stress-induced PK cascades to complex environmental stimuli [14]. In higher plants, PP2Cs negatively regulates the ABA signaling pathway and decrease tolerance against oxidative stress [15]. Proteins encoded by PP2Cs candidate genes play a vital role in various abiotic stress signaling such as Salt, drought, and freezing [16]. PP2C family has been identified and characterized in Arabidopsis, rice maize, and cotton among species as such developed complex molecular mechanisms to implement and survive during adverse growth conditions [17, 18]. These studies have established the diverse role of PP2C genes in plant development and environmental stresses. However, not conducted in *Corchorus capsularis* despite its' utilization and productivity have been affected due to salt stress. Therefore, this study is crucially needed and urgent. However, genome data availability allows us to conduct this study. Twenty-four transcriptome sequencing from the root and leaf tissues of salt-tolerant and sensitive jute germplasms was used to identify potential salt-stress responsive genes and clarify regulatory pathways involved in salt tolerance jute. Moreover, comprehensive analysis of *Corchorus capsularis* PP2C regards; the phylogenetic relationship between *Corchorus capsularis* and Arabidopsis PP2C, conserved motifs, gene structure, the expression profile of the key stress marker genes (RAB18, RD29B, KIN1, and RD29A), as well as chromosome mapping of *Corchorus capsularis* PP2C genes were also investigated.

# Results

## Transcriptome assembly and sequencing

Transcriptome analysis of the twenty-four Salt tolerant (J194) and sensitive (J7) samples using Illumina paired-ends sequencing technology was used to explore the DEGs related to NaCl stress in jute (utilizing two biological replicates in each case; tissue (leaf and roots) and time (0, 6 and 12 hours) (Additional Table S3). The comparison of J194 and J7 due to salt exposure is also presented in Figure (1a-b). The 1518813542 raw reads were generated, 759406771 clean reads were obtained, and the sequence was mapped to the jute genome (Additional Table S3). The GC contents range between 44.24–45.59%, and the Q30 scores are greater than 90% (Additional Table S3 and S4). The clean reads were used to assemble the 27720 transcripts (Additional Table S3) to be used as a reference sequence for downstream analysis and 26463 unigenes (Additional Table S4) 7272 unigenes. Comparing the clean reads with the reference genome when assigned into exon, intron, and intergenic regions indicated that the exon-region sequence accounts for more than 80% of the genome-mapped sequences as expected (Additional Figure S1). These results present a higher degree of annotation accuracy. Additionally, the box plot comparison of Fragments Per Kilobase of transcript per Million mapped reads (FPK) indicates that sequence results were reliable; the sample yielded equivalent reads and coverage depths between duplicates (Fig. 1b).

To comprehend the DEGs annotated unigenes functions, 26463 unigenes were annotated in all eight databases. It was observed that most of the unigenes (26413) were aligned with that of the NR database, followed by eggNOG\_Annotation (22539) and Pfam\_Annotation (19493). However, annotation in the KEGG database (8992 unigenes) presented the least number of unigenes (Fig. 1c). DEGs annotations across the treated samples (Additional Table S5) indicated that (J7R\_0H Vs. J7R\_6H) had the highest number of annotated unigenes across all the databases, followed by (J7L\_0H Vs. J7L\_6H), whereas (J194L\_0H Vs. J194L\_6H) had the least number across all the databases (Additional Table S5). It was observed that the most common enrichment analysis was related to plant hormone signal transduction pathways (Additional Table S6) as vital for jute response to NaCl exposure.

## Identification of differentially expressed genes in response to salt treatment

Analysis of DEGs in J7 and J194 on different tissues (roots and leaves) under control and salt stress conditions. Indicated that 26463 unigenes were differentially expressed (Additional Table S7). Cluster analysis revealed that differential gene expression profiles were similar within the same tissue types and time exposure rather than between different genotypes and time duration (Fig. 2a). The dispersion of these differentially expressed genes is presented in (Additional Figure S2).

Additional Figure S2 (i) indicated the same germplasm combination for up-regulated and down-regulated DEGs in J7 and J194 leaf tissues at all durations of NaCl exposure. Up-regulated DEGs (Figure c) indicated no DEGs were recorded across all the germplasms in all durations. However, J7\_0H Vs. J7\_6H (1075) and J194\_0H Vs. J194\_6H (6) recorded the highest and lowest DEGs, respectively. Moreover, (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) (1052) had a high combination, but none were observed (J194-0H Vs. J194-6H) Vs. J7-0H Vs. J7\_12H) (0). However, 16 and 2 DEGs were recorded in (((J7\_0H Vs J7\_6H) Vs (J7\_0H Vs

J7\_12H) Vs (J194\_0H Vs J194\_12H))) and (((J7\_0H Vs J7\_6H) Vs (J194\_0H Vs J194\_12H) Vs (J194\_0H Vs J194\_6H))) combinations, respectively. Figure (a) indicated the down-regulated DEGs for the same germplasm combination across all durations of exposure to NaCl indicated that 11 DEGs were recorded across all the germplasms and the time of the exposure to salt treatment. J7\_0H Vs. J7\_6H (840) and J194\_0H Vs. J194\_6H (6) had the highest and lowest number of DEGs. Still, (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) (1093) recorded the highest DEGs, whereas none was observed in (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) (0). 43 and 4 DEGs were recorded in (((J7\_0H Vs J7\_6H) Vs (J7\_0H Vs J7\_12H) Vs (J194\_0H Vs J194\_12H))) and, (((J7\_0H Vs J7\_6H) Vs (J7\_0H Vs J7\_12H) Vs (J194\_0H Vs J194\_6H))) combinations, respectively. Moreover, figure b and d indicated the DEGs under different germplasms combined for down and up-regulated leaf tissue, indicating that about 33 and 21 DEGs were recorded in (J194\_6H Vs. J7\_6H) Vs. (J194\_12H Vs. J7\_12H) and (J194\_6H Vs. J7\_6H) Vs. (J194\_12H Vs. J7\_12H) respectively.

Additional Figure S2 (ii) indicated different germplasm combination for up and down-regulated DEGs in J7 and J194 root tissues at all durations of exposure to NaCl. For Up-regulated DEGs (Figure c), indicated that 67 DEGs were recorded in all the germplasms across the duration of the exposure to NaCl. Still, J7\_0H Vs. J7\_6H (1380) and J194\_0H Vs. J194\_6H (50) recorded the highest and lowest number of DEGs, respectively. Interestingly, (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) (168) and (J194\_0H Vs. J194\_12H) Vs. (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) (4) possess the highest and lowest number of DEGs. Hitherto, 291 and 9 DEGs were recorded for (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) and (J194\_0H Vs. J194\_6H) Vs. (J194\_0H Vs. J194\_12H) combinations, respectively. J7 and J194 down-regulated DEGs roots are presented in Figure (a), in which 258 DEGs were recorded across all germplasms at a different time of NaCl exposure J7\_0H Vs. J7\_6H(1044) and J194\_0H Vs. J194\_12H (38) had the highest and lowest numbers. Additionally, (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) (550) and (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) (6) recorded the highest and lowest DEGs combination, respectively. Moreover, 218 and 8 DEGs were recorded in (J7\_0H Vs. J7\_6H) Vs. (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) and (J194\_0H Vs. J194\_12H) Vs. (J7\_0H Vs. J7\_12H) Vs. (J194\_0H Vs. J194\_6H) combinations, respectively. Also, (Figures b and d) indicated the DEGs under different germplasms combination for down and up-regulated roots tissue, respectively; the result revealed that about 16 and 12 DEGs were observed in (J194\_6H Vs. J7\_6H) Vs. (J194\_12H Vs. J7\_12H) and (J194\_6H Vs. J7\_6H) Vs. (J194\_12H Vs. J7\_12H) respectively.

Comparison of Up/down-regulated DEGs for J7 and J194 leaves and roots at 0, 6, and 12 hours of exposure to NaCl are presented in (Fig. 2b). The results indicated that J7R\_0H Vs. J7R\_6H (2413) had the highest number of up-regulated DEGs, followed by J7L\_0H Vs. J7L\_6H (2152), whereas J194L\_0H Vs. J194L\_12H (38) had the lowest number. However, J7L\_0H Vs. J7L\_6H (2256) had the highest number of down-regulated DEGs, followed by J7L\_0H Vs. J7L\_6H (2000) whereas J194L\_0H Vs. J7L\_0H (74) recorded the least number. However, J194R\_0H Vs. J194R\_6H recorded up and down-regulated DEGs of 703, 689 respectively.

Cluster analysis revealed that differential gene expressions were more similar within the same tissue types and time exposure than between different genotypes and time duration (Fig. 2a). We observed that genes

within the same tissues under the exact period of exposure to NaCl were clustered together. Additionally, replicating the treated and control samples was highly consistent, signifying the RNA-seq results' reproducibility.

The results showed that plant hormone signal transduction (ko04075) was the most enriched pathway in root tissues at six-hour exposures to NaCl stress. Plant hormones such as auxins, jasmonic acid, and abscisic acid play significant roles in the plant response to stresses by regulating hormone signal transduction pathways [19, 20]. Among the 62 enriched DEGs observed in both J194 and J7 root tissues at six hours of exposure to NaCl Stress (Additional Table S8), 17 had unknown functions, while the remaining 45 were generally related to abiotic stress responses.

Plant hormonal signal transduction pathways of the up-and down-regulated DEGs in the two germplasms' root tissues (J194 and J7) at six-hour of exposure to NaCl stress are presented (Fig. 2c and Additional Figure S3). The result indicated that three DEGs were involved for AUXIN1, including *Corchorus\_capsularis\_newGene\_739*, Cc.01G0031930, and Cc.04G0019670. These genes were down-regulated in J194 and J7, respectively. For IAA, five DEGs were identified; however, only Cc.07G0002340 (down-regulated) and Cc.07G0003430 (up-regulated) were recorded in J194. Additionally, two DEGs (Cc.01G0010780 and Cc.07G0003430) and three DEGs (Cc.04G0043620, Cc.04G0045370 and Cc.07G0002340) were up-and down-regulated in J7, respectively (Fig. 2c). For ARF, three DEGs were involved, but only one (Cc.06G0025590) was down-regulated in both J194 and J7. another one (Cc.06G0027580) was down-regulated in J7 only. For IAA, one DEG (Cc.07G0005110) was up-regulated in both J194 and J7, While for the GH3 signaling pathway, two up-regulated DEGs were found (Cc.01G0028710 and Cc.02G0006030) in both J194 and J7. SMALL AUXIN UP RNAs (SAUR) involves eight DEGs, among which four (Cc.03G0019440, Cc.03G0029630, Cc.04G0007220 and Cc.05G0005100) were up-regulated in both J194 and J7. However, the remaining four (Cc.02G0011530, Cc.02G0022670, Cc.04G0014260, and Cc.06G0027410) were up-regulated only in J7.

For PYL, one DEG (Cc.03G0016680) was only up-regulated in J7 whereas six DEGs were involved in the PP2C pathway, out of which three (Cc.03G0000600, Cc.03G0030800, and Cc.07G0001880) were up-regulated in both J194 and J7, Cc.03G0016550, and Cc.07G0028160 were up-regulated in J194 while Cc.06G0030850) was only up-regulated in J7. SnRK2 signaling pathway had only three DEGs, including Cc.02G0003620 and Cc.04G0017920 that were up-regulated, and Cc.02G0021190, which was down-regulated in J7. For ABA-responsive element binding factor (ABF), three DEGs were involved; one (Cc.04G0004780) was up-regulated in both J194 and J7, whereas Cc.01G0035870 and Cc.06G0010680 were respectively up-and down-regulated specifically in J7. Two DEGs were involved in JASMONATE ZIM DOMAIN (JAZ) pathway; however, only one (Cc.06G0030170) was up-regulated in both J194 and J7. In the MYC2 pathway, only one gene (Cc.04G0013900) was down-regulated in both J194 and J7.

Quantitative reverse transcription-PCR to validate the RNA-seq result is presented in (Additional Figure S4). The result indicated that all the study genes confirmed their expression in both the qRT –PCR and RNA-seq, except that sample (J7R\_12H Vs. CR) for ABF/bzip genes were down-regulated in the qRT –PCR rather than up-regulated.

Correlation analysis between germination rate related traits with seedling stage parameters and qRT-PCR with RNA-seq results are presented in supplementary (Additional Figure S5 and S6) respectively. The results indicated a significant positive correlation between germination related traits (RGS and RGVS) and seedling stage parameters exposed to NaCl stress, except DRW, RL, RFW, and DSW. Moreover, a significant positive correlation was also observed between qRT-PCR and RNA-seq results (Additional Figure S6).

The relative expression levels of seven randomly selected candidate genes in J194 and J7 at one and two weeks duration of exposure to NaCl stress are presented in (Additional Figure S7a-b) validating the candidate genes. The gene relative expression levels of the J194 and J7 at one week is shown in (Additional Figure S7a and c). The relative expression of the genes in leaf and root tissues of J194 and J7, were higher in J194 compared to J7, though the expression of Cc.06G0024090 (MYB) was higher in the control condition than that of the treated samples of both J194 and J7. The gene relative expression levels of the J194 and J7 at two weeks is presented in (Additional Figure S7b and d). The result indicated that the expression levels were slightly higher than that of one week. Still, the genes expressed high in treated J194 root and leaf tissue samples compared to that of J7; however, Cc.06G0024090 (MYB) of the control leaf sample had the highest expression compare to the treated samples. It was observed that the genes expressed more elevated in the leaf than in the root.

### **Identification of *Corchorus capsularis* PP2C gene family**

The *Corchorus capsularis* PP2C gene family and their homologs in other species are presented in (Table 1); the table indicated 89.8–100% relative similarities and most of the homologous organisms were *Theobroma cacao* (15 genes), *Herrania umbratica* (8 genes), and *Duriozibethinus* (5 genes). Moreover, the identified PP2C Arabidopsis and *Corchorus capsularis* conserved domains' after removing redundancies sequences, and non-PP2C using SMART and PFAM (Additional Table S10). About 78 Arabidopsis PP2C were identified, which used as a queries sequences to search against the *Corchorus capsularis* genome using NCBI local blast as stated in the materials and methods, after confirming the presence of PP2C catalytic domains using SMART and PFAM, about 38 genes were identified (Additional Table S10).

To study *Corchorus capsularis* and Arabidopsis and PP2C protein families' relationship, we generated an alignment for full-length protein sequences of 76 and 38, respectively (Fig. 3). To be consistent with the previous report of Arabidopsis PP2C [21], we followed the same name for twelve common subfamilies (A, B, C, D, E, F (1 and 2), G, H, I, J, K, and L) [21]. Moreover, three novel groups M (AT2G40860 and Cc.05G0014550), N (AT4G27800 and Cc.03G0020320), and O (AT5G19280 and Cc.04G0042080) were identified with 100% bootstrap support. Group F, as previously reported, and was divided into two sub-groups (F1 and F2) with 99 and 100% bootstrap support, indicating that these genes in the same sub-family shared a similar evolutionary origin. We also observed that in-group F2 (AT5G53140 and Cc.03G0016550) as well as (AT4G31750 and Cc.07G0024730) shared 100% bootstrap, revealing they do share an evolutionary relationship. However, F1, B, and K groups of Arabidopsis PP2C proteins did not share similarities' with *Corchorus capsularis* PP2C. In-group A, Cc.06G0030850 had 57% bootstrap support with AT1G17550, ATG72770, AT4G26080, and AT5G57050. Moreover, in-group G, Cc.07G0010300 had 100% bootstrap similarities with AT2G25620, Cc.01G0009520 had 100% bootstrap support with

AT3G51570 and AT2G33700. Cc.06G0018350 in-group I had 100% bootstrap support with AT2G25070 and AT4G31860. Group M and N, Cc.05G0014550 and Cc.03G0020320 had 100% bootstrap support with AT2G40860 and AT4G27800. Group H, Cc.040004590 and Cc.05G0029950 had 100% bootstrap support respectively. Additionally, Cc.07G0002210 and Cc.04G0045880 had 100% similarities respectively. For group E, Cc.05G0029870 had 100% bootstrap support with AT5G27930 and AT3G05640, however 66% with Cc.04G0047530. Additionally, in this group, Cc.06G001817 had 100% bootstrap with AT1G16220 and AT1G79630; lastly, AT5G36250 and AT3G02750 had 100% and 99% bootstraps support respectively with Cc.02G0024270. Furthermore, in the J group, Cc.04G0009430 had 100% bootstrap with AT3G63320 and AT3G63340. Regarding group C, Cc.05G0002960 and Cc.07G0028160 had 99 and 100% bootstrap support with (ATG07630, AT2G28890, AT3G09400, and AT5G02400) and (ATG07630 and AT2G28890), respectively. Moreover, AT2G46920, AT2G35350 and AT3G16560 had 100% bootstrap support with Cc.0440013340, Cc.03G0022250 and Cc.04G0046980 respectively. In-group D, Cc.05G0010230 and Cc.04G0048240 had 100% bootstrap support with AT5G06750 and AT3G17090 respectively. Moreover, AT4G33920 had 100, 52% bootstrap similarities with Cc.02G0005900 and Cc.01G0028490 respectively, furthermore Cc.03G0003110 had 100 and 97% similarities with AT3G55050 and AT3G12620 respectively. Similarly, Cc.01G0032360 and Cc.02G0000770 had 73% bootstrap with AT3G51370 and AT5G66080. Group O Cc.04G0042080 had 100% bootstrap support with AT5G19280.

Table 1  
*Corchorus capsularis* PP2C gene family and theirs homologs in other species

Similar Jute gene bank ID					Homologous genes
Gene bank ID	Gene bank ID	RS(%)	Gene length(bp)	Protein(aa)	Gene bank ID
Cc.01G0009520	OM083396.1	100	1074 bp	357 aa	XP_021284306.1(Herrania umbratica)
Cc.01G0028490	OM073934.1	100	112028 bp	381 aa	XP_021276140.1(Herrania umbratica)
Cc.01G0032360	OM074554.1	100	48004 bp	397 aa	XP_007041776.2[Theobroma cacao]
Cc.02G0000770	OM055759.1	100	96995 bp	397 aa	XP_007046602.1[Theobroma cacao]
Cc.02G0005900	OM056534.1	100	46132 bp	380 aa	XP_022751230.1[Durio zibethinus]
Cc.02G0024270	OM052614.1	96.02	26475 bp	493 aa	XP_021282751.1[Herrania umbratica]
Cc.03G0003110	OMP08102.1	97.31	38078 bp	331 aa	XP_007009638.1[Theobroma cacao]
Cc.03G0016650	OM099628.1	100	34627 bp	354 aa	XP_021295755.1[Herrania umbratica]
Cc.03G0020320	OM087933.1	100	85758 bp	392 aa	XP_022734118.1[Durio zibethinus]
Cc.03G0022250	OM077235.1	100	50299 bp	789 aa	XP_007039547.2[Theobroma cacao]
Cc.04G0004590	OM078977.1	100	71295 bp	428 aa	XP_017981229.1[Theobroma cacao]
Cc.04G0009430	OM077631.1	99	86720 bp	1088 aa	EOX94652.1 [Theobroma cacao]
Cc.04G0013340	OM082327.1	89.8	36529 bp	896 aa	EOX95490.1[Theobroma cacao]
Cc.04G0042080	OM061226.1	100	169640 bp	590 aa	XP_022762030 [Durio zibethinus]
Cc.04G0045880	OM098359.1	98.59	67103 bp	426 aa	XP_021290662.1[Herrania umbratica]
Cc.04G0046980	OM053660.1	100	151752 bp	531 aa	XP_022760559.1[Durio zibethinus]
Cc.04G0047530	OMP11874.1	95.09	14476 bp	410 aa	XP_021290531.1 [Herrania umbratica]

Similar Jute gene bank ID					Homologous genes
Cc.04G0048240	OM053492.1	99.41	22592 bp	1173 aa	XP_021291718.1[Herrania umbratica]
Cc.05G0002960	OM066259.1	100	49282 bp	597 aa	XP_007027029.2[Theobroma cacao]
Cc.05G0010230	OM063000.1	100	122420 bp	388 aa	XP_017977120.1 [Theobroma cacao]
Cc.05G0014550	OM053906.1	97.42	58784 bp	659 aa	XP_022742528.1[Durio zibethinus]
Cc.05G0029870	OMP11453.1	99.73	101706 bp	369 aa	XP_017977558.1[Theobroma cacao]
Cc.06G0018170	OM083325.1	100	15947 bp	497 aa	EOY26142.1[Theobroma cacao]
Cc.06G0018350	OM074037.1	98.9	106105 bp	362 aa	XP_021298273 [Herrania umbratica]
Cc.06G0030850	OM070679.1	100	96546 bp	539 aa	XP_007024551.2 [Theobroma cacao]
Cc.07G0002210	OM082641.1	100	103588 bp	1085 aa	XP_017983109.1[Theobroma cacao]
Cc.07G0010300	OM080598.1	93.68	58017 bp	364 aa	XP_007013785.1[Theobroma cacao]
Cc.07G0028790	OM067773.1	98.76	53567 bp	733 aa	XP_007016425.2[Theobroma cacao]

The distance tree was inferred using the neighbor-joining methods based upon an alignment of full-length amino-acid sequences of the PP2C conserved domains. They were grouped according to the PP2C Arabidopsis genes previously reported.

For identification of motifs and gene structure, 11 *Corchorus capsularis* PP2C motifs were identified; their sequences and motifs distributions are presented in (Fig. 4a). The results indicated that motifs 3, 2,5,11, 4, and 1 were wide spreads among the *Corchorus capsularis* PP2C groups. However, other motifs such as 9 and 6 were specific to two groups (D and C), whereas 10 and 7 were exclusively specific to group D. Moreover, the 8 motif was specific to group C and E.

The phylogenetic tree was constructed from alignments of 38 amino acid sequences of PP2C conserved domains in *Corchorus capsularis* PP2C; it was grouped according to PP2C Arabidopsis genes previously reported. The Sequences and the length of the conserved motifs in the amino acid sequences of *Corchorus capsularis* PP2C genes are also indicated at the bottom.

The arrangement of exon-intron will play an important role in the process of *Corchorus capsularis* PP2C diversification. As such, exon-intron organization was analyzed, using online tools GSDS as stated in the methodology. Figure 4b illustrated the distribution of the *Corchorus capsularis* PP2C exon-intron; the result revealed that most of the *Corchorus capsularis* PP2C members in the same subfamily shared similar exon number with different exon and intron length, the number of exon-intron ranges between (3–21) and (2–20), respectively. Cc.04G0046980 had the least number of exon (3), and intron (2) subfamily group C. Whereas Cc.04G0048240 in subfamily member of group D had the highest exon and intron number of 20 and 19 respectively. Group G and A had four exons, except for Cc.06G0018350 of group G, which had ten. Cc.03G0016550 and Cc.07G0024730 of group F2 had a similar number (8) of the exon. Cc.05G0029870 and Cc.04G0047530 of group E had the same number (5) of the exon. Cc.05G0002960, Cc.07G0028160 and Cc.03G0022250 of group C had the same number (4) of exon. Moreover, group D member (Cc.02G0005900, Cc.05G0010230, Cc.03G0003110, Cc.01G0032360, and Cc.02G0060770) had the same number (4) of exon.

The phylogenetic tree was constructed from alignments of 38 amino acid sequences of PP2C conserved domains in *Corchorus capsularis* PP2C; they were grouped according to previously reported Arabidopsis PP2C. The exon/intron structure of each *Corchorus capsularis* PP2C gene was proportionally displayed according to the scale at the bottom. Yellow boxes represent exons, gray lines represent introns and blue boxes represent untranslated regions.

The result further indicated that *Corchorus capsularis* PP2C genes were localized on 7 chromosome groups, and were unevenly distributed (Additional Figure S8). Three genes each were mapped on chromosome 1 and 2, however about 4 genes were mapped on chromosome 7. Moreover, chromosome 4 and 5 had the highest number of map genes (Cc.04G0004590, Cc.04G0009430, Cc.04G0013340, Cc.04G0042080, Cc.04G0046980, Cc.04G0048240, Cc.04G0045880, and Cc.04G0047530) and (Cc.05G0010230, Cc.05G0029870, Cc.05G0002960, Cc.05G0014550, and Cc.05G0029950), respectively.

### **Expression pattern of PP2C (reference marker) genes in J194 and J7**

Response to NaCl stress was assessed by monitoring the expression pattern of the key stress marker genes in J194 and J7 under control and stress conditions, as mentioned in the materials and methods. As such, we conducted qRT-PCR base expression analysis from the well-established stress marker genes such as; RAB18, RD29B, KIN1, and RD29A using J194 and J7 root and leaf tissue samples exposed to NaCl at 0, 6, and 12 hours as previously stated. Besides, J194 and J7 root and leaf tissue samples at one and two weeks' of exposure to NaCl were conducted.

The relative gene expression level of Cc.04G002237(RAB18), Cc.03G0029910(RD29B), Cc.01G0008270(KIN1), and Cc.03G0029910(RD29A) at 0, 6, and 12 hours of exposure to NaCl are presented in (Figs. 5a-d). The result indicated salt-tolerant germplasm (J194) had the highest relative gene expression level across the periods. The result further revealed that J194 root tissues at six-hour exposure to NaCl recorded the highest gene expression levels in all the studied genes.

Moreover, these genes' relative expression levels at one and two-week exposure to NaCl are presented (Figs. 6a-d). The results still indicated the J194 had the highest relative gene expression level in all studied genes across the periods of exposure to NaCl. However, the relative expression level of Cc.04G002237 (RAB18) and Cc.03G0029910 (RD29B) were higher in J194 roots tissues across the period of NaCl exposure whereas, Cc.01G0008270 (KIN1) and Cc.03G0029910 (RD29A) had the highest expression level in J194 leaf tissues at all duration of exposure to NaCl.

The qRT-PCR analysis was conducted to generate the relative expression profile level of key established stress marker genes such as RAB18, RD29B, and KIN, and RD29A (a to d respectively) in J194 and J7 leaf (L) and root (R) tissues at CT(0), 6, and 12 hours as indicated in X-axis and the relative expression level is indicated in Y-axis. Data from the mean of the replicated samples are presented as columns and an error bar denotes the standard deviation. \*p-value < 0.05 and \*\*p-value < 0.01 indicate statistically significant level.

The qPCR analysis was conducted to generate the relative expression profile level of key established stress marker genes such as RAB18, RD29B, and KIN, and RD29A (a to d, respectively) in J194 and J7 leaf (L) and root (R) tissues at 0 (control), one, and two weeks duration of the exposure to NaCl. The X-axis represents the samples at control (C) and Treated (T) and the relative expression level is indicated in Y-axis. Data from the mean of the replicated samples are presented as columns and an error bar denotes the standard deviation. \*p-value < 0.05 and \*\*p-value < 0.01 indicate statistically significant level.

## Discussions

Identification of genes and their functions can be achieved by clustering them into the same or similar expression patterns using hierarchical clustering analysis based on their expression levels determined from FPKM values. In general, those genes' metabolic patterns with the same/similar functions or functioning in the same pathways are clustered together. In this study, we observed that genes from the same tissues exposed to NaCl at the same duration clustered together (Fig. 2a). DEGs within a single cluster appeared as co-expressed genes. In contrast, color-coding of different cluster groupings signifies genes with similar expression patterns that share the same functions, thus participate in the same biological processes.

Furthermore, we observed that for up- and down-regulated DEGs in leaf tissues, J7 at six and twelve-hour of exposure to NaCl as well as their combinations had the highest number of DEGs (Additional Figure S2(i)). Additionally, J194, at six and twelve hours exposure to NaCl and their pairing, had the least number of DEGs. The number of DEGs across the two germplasms (J7 and J194) and the duration of exposure to NaCl in root tissue indicated that J7 at six and twelve-hour of exposure to NaCl as well as their combinations had the highest number of DEGs (Additional Figure S2(ii)), the numbers are much higher at six hours than at 12 hours. Additionally, still J194 at six and twelve hours of exposure to NaCl as well as their combinations had the least number of DEGs. The results are consistent with several previous studies [22–24]. Indicating that salt-sensitive germplasms display the highest number of DEGs than tolerant germplasms, and 6 hours of exposure to NaCl recorded the highest number of DEGs than 12h.

Moreover, this is because the effects of salt stress on the salt-sensitive germplasm was greater than that of tolerant germplasm. Interestingly, we recorded a higher number of DEGs after pairing the sensitive germplasms (J7) at 6 and 12 hours combination than pairing the tolerant germplasms (J194) (Figures S2(I-II)). From the results, we observed that J194 (salt-tolerant) had fewer DEGs than J7 (salt-sensitive) (Additional Figures S2 (I-II)). These findings agree with those reported from previous studies [22–24]. Moreover, the roots of J194 and J7 had more DEGs than the leaves that indicates there are other genes associated with roots that do not necessarily play roles in plant response to salt stresses.

The ABA signaling pathway comprises many regulated genes and plays a role in plant response to stress. It also mediates the mechanism for adaption to environmental changes [6]. Usually, SnRK2 positively regulates ABA signaling, critical in abiotic stress responses, especially NaCl stress. A high level of ABA during the stages of seedling growth is maintained, from which the plants are protected from damage caused by salt stress [25]. The plant hormone auxin plays a vital role in regulating plant developmental processes [26]. In the present study, most of the auxin-related genes observed in the root tissues exhibited similar expression patterns in J7 and J194 samples at 6 hours post-exposure to NaCl, which indicates their significance in plant abiotic stress responses and conserved mechanism of action. Auxin carriers such as the auxin influx carrier 1 mediate auxin's polar transport (AUX1) [27]. It was observed that hormone signaling pathways were explicitly up-regulated in J194 and J7, reflecting their significant plant stress tolerance roles. Plant hormone signal transduction pathways were the most enriched. About seven identified novel DEGs, one AUXIN1 DEG (*Corchorus capsularis*\_newGene\_739) was down-regulated in both the salt-tolerant and sensitive germplasms in root tissues at 6 hours post-exposure to NaCl.

For the PYL-ABA-PP2C pathway, we recorded seventeen DEGs (Cc.03G0016680, Cc.03G0000600, Cc.03G0016550, Cc.03G0027770, Cc.03G0030800, Cc.06G0030850, Cc.07G0001880, Cc.07G0028160, Cc.07G0031700, Cc.02G0003620, Cc.02G0021190, Cc.03G0023450, Cc.04G0017920, Cc.07G0021650, Cc.01G0015540, Cc.01G0035870 and Cc.04G0004780) in the salt-stressed root tissues of both J194 and J7 which is consistent with some previous studies [24, 28]. Interestingly, our results showed that the PYL gene (Cc.03G0016680) was up-regulated, supporting the basic ABA signaling model but contradicts the findings [24]. Under normal circumstances, PYL binds to ABA and PP2Cs to form PYL-ABA-PP2C complexes, thereby inhibiting PP2Cs. This inhibition releases autophosphorylating SnRK2s, which then phosphorylate many downstream effectors [3]. As such, PYL expression is anticipated to keep pace with ABA concentration and be up-regulated under stress conditions.

Additionally, Our KEGG analysis confirmed the enrichment of plant hormone signal transduction (Additional Table S9 and Fig. 2c) in the root tissues at 6 hours of exposure to NaCl. By regulating hormone signal transduction pathways, plant hormones such as auxins, jasmonic acid, and abscisic acid play a significant role in plant response to abiotic stresses [19, 20]. The importance of ABA in plant responses to these stresses is conserved. Interestingly, from the results, plant hormone signal transduction pathway (Fig. 2c), three *Corchorus capsularis* PP2C genes (Cc.06G0030850, Cc.03G0016550, and Cc.07G0028160) were found to be up-regulated in J194 root at 6 hours exposure to NaCl, moreover have 57, 100 and 99% bootstrap support with Arabidopsis PP2C (AT1G17550, AT1G72770, AT4G26080, and AT5G57050), AT5G53140, and (AT1G07630 and AT2G28890) genes respectively. Also, the chromosomal location of such

genes (Additional Figure S8) indicated their involvement in segmental duplication. As such, these genes could be serves as a candidature for salt tolerance in *Corchorus capsularis*.

Moreover, phytohormone abscisic acid plays vital regulatory roles in salt, drought, and cold stresses during the plant developmental stages [29, 30]. As such, NaCl stress was assessed by monitoring the expression pattern of the key stress marker genes (RAB18, RD29B, KIN1, and RD29A) in J194 and J7 under control and stress conditions. Our results indicated that the expression of these stress marker genes in salt-tolerant individuals (J194) was higher than the sensitive individual (J7) at six hours of exposure to NaCl in root tissues. Our findings were consistent with those reported by Singh et al.[31], which signifies that the over-expression profiles of *Corchorus capsularis* PP2C (Cc.04G002237, Cc.03G0029910, Cc.01G0008270, and Cc.03G0029910) genes revealed a significant interplay of ABA-dependent and independent pathways for abiotic stress. However, at one and two weeks of exposure to NaCl, still indicated J194 had the highest gene expression levels. Yet, the expression profile of these stress marker genes reveals that they were tissue-specific. Indicated that RAB18 and RD29B were highly expressed in root tissues, whereas KIN1 and RD29A expressed highest in leaf.

The analysis of gene structure and conserved motifs based upon the phylogenetic relationship was conducted to get insight into the structural feature relationship of the *Corchorus capsularis* PP2C. Eleven number of Motifs were identified, which is consistent with those reported by Cao et al.[14, 21] and different from the one stated in cotton [32]. Must of the PP2C family members contained 3, 2, 5,11,1, and 4 motifs except for group D that had some specific motifs (7, 10, 9, and 6). The results revealed that all the identified PP2C genes have some evolutionary relationship; thus, proteins classified into the same subgroup shared the same sequence. For *Corchorus capsularis* gene structure PP2C, the number of exon-intron ranges between 3–21, and 2–20, respectively that contradicted the findings of Cao et al. [14, 21, 32]. These results revealed that PP2C genes in the same subgroup show more or less similar exon-intron organization. The work further revealed that the differences in motifs distribution in respective protein sequences provide divergence evidence of gene functions in different sub-families. Moreover, multiple alignments revealed that not all *Corcharus capsularis* PP2C domains contain all the conserved motifs due to the partial deletion in the c-terminal of the PP2C phosphate catalytic domains. As such, resulting in a few motifs and loss of functions, similar findings were reported [14]. However, proteins in the same subfamily exhibit equal motif distribution to support their close evolutionary relationship. Surprisingly, indicated in (Figs. 3 and 4b) the *Corchorus capsularis* PP2C proteins remain in the same subfamily due to the pattern of the *Corchorus capsularis* PP2C conserved domains exon-intron gene structure were examined according to their phylogenetic relationship. Moreover, the protein domains and exon-intron structural diversity pattern play a vital role in gene family evolution; these findings were consistent with those reported by Cao et al.[14, 32]. Moreover, we identified 78 Arabidopsis PP2C, in contrast to 76 and 80 identified by Kerk et al.[17, 21], respectively. This novel Arabidopsis PP2C observed in groups M (AT2G40860), N (AT4G27800), and O (AT5G19280) had 100% bootstrap support with Cc.05G0014550, Cc.03G0020320, and Cc.04G0042080 of *Corchorus capsularis* PP2C, respectively.

Small-scale tandem and large segmental duplications are the two central mechanisms contributing to the plant kingdom's genome complexities [33]. Xue et al.[21], reported that PP2C expanded through

chromosomal duplication and the whole genome. Moreover, as mentioned by Cheung et al.[34], closely related genes located within a distance of fewer than 200 kb on the same chromosome are referred to as tandem duplications; otherwise, segmental duplications. Two pair's paralogous *Corchorus capsularis* PP2C genes on chromosome Five (Cc.05G0029950 and Cc.05G0029870) and Four (Cc.04G0047530 and Cc.04G0048240) were found to be involved in tandem duplication events others were segmental duplication. As such, most of the PP2C *Corchorus capsularis* genes were involved in segmental duplication, and this is consistent with the findings of [14, 32]. This duplication type plays a vital role in expanding the PP2C *Corchorus capsularis* gene family. It might have experienced functional discrepancies and probably may lose its unique functions, rather than developed novel functions [35].

## Conclusion

Our findings provide detailed transcriptome analyses of salt-tolerant and salt-sensitive jute germplasms. Plant hormones such as auxins, jasmonic acid, and abscisic acid play significant roles in the plant response to stresses by regulating hormone signal transduction pathways, indicating ABA's importance in the plant responses' stresses conserved. We found 38 *Corchorus capsularis* PP2C, comparative phylogenetic analysis of *Corchorus capsularis*, and *Arabidopsis* PP2C categorized PP2C into 15 based upon conserved domains. Eleven conserved motifs were identified, and most of the genes had the same number. The exon-intron ranges between (3–21) and (2–20), respectively. Most of the *Corchorus capsularis* PP2C genes were involved from segmental duplication, and analysis of the key stress marker salt-tolerant PP2C genes validated the salt tolerance individuals. These results provided valuable insight into salt tolerance transcriptome and indicated that PP2C had provided a stepping-stone to the molecular mechanism in *Corchoruscapsularis*. Furthermore, differentially expressed genes, motifs, gene structure, and the chromosomal location of salt tolerance candidate genes might have experienced functional divergence. As such, their further study will enhance salt tolerance in *Corchorus capsularis*.

## Materials And Methods

### Source of plant materials

About 292 jute germplasms were used for the screening of tolerant and sensitive individuals at the germination stage. The preliminary germination experiment was set-up to select suitable NaCl concentration for the study. Different NaCl concentrations (40mM, 120mM, and 200mM) and the Control (Deionized water) were tested on the parent strains. Petri dishes of about (12 x 12 x 5) cm with layers of filter paper were used to germinate 50 seeds from each parent. 18mls of the prepared treatment solutions (NaCl dissolved in deionized water) and control (deionized water) was poured into the Petri dishes, and the seeds were subsequently spread accordingly. A Completely Randomized Design was used and repeated three times. Petri dishes were kept in a growing chamber and maintained at 28<sup>0</sup>C during the day and 22<sup>0</sup>C at night on 16-hour light per 8-hour dark cycle for six days. Germinated seeds were recorded as germination percentage (number of germinated seed/total number of seeds sowed) x100. Based upon the mean germination percentages of the two parents generated (result not shown), suitable concentrations of NaCl were identified as 120mM NaCl (result not shown), which was used to screen out the 300 germplasms in

response to NaCl concentration, following the same experimental protocol. Two germplasms, salt-tolerant (J194) and salt-sensitive (J7) were selected for transcriptomic analysis. They were further subjected to different salt concentrations (200, 250, and 300mM) at the seedling stage to obtain a suitable concentration for the experiment using Pindrustup media. The salt solutions were applied at two to the three-leaf stage for two weeks. Data were taken and analyzed, based upon which 250mM concentration was selected (result not shown).

### **Salt treatment and sample preparations**

J194 and J7 were grown on a separate pot filled with Pindrustup media under normal conditions. At two to three leaves stage, seedlings from each sample (J194 and J7) were rinsed with distilled water and immediately placed into the tubes filled with 250mM NaCl solution and control (distilled water (CK)). Root and leaf tissues from each germplasm were harvested, frozen immediately in liquid nitrogen, and then stored at -80°C for RNA extraction. Two biological replicates of each sample were used for RNA extraction.

### **RNA-extraction and isolation**

Total RNA was isolated from the leaf and root tissue samples using Trizol (Invitrogen, Santa Clara, CA, USA), following the manufacturer's protocol. RNA degradation and contamination were checked in 1% agarose gels. RNA purity and concentration were checked using NanoPhotometer® spectrophotometer (Implen, West Lake Village, CA, USA) and Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), respectively. The RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). Twenty-four samples were used for the transcriptome sequencing of J194 and J7 roots and leaf tissues at 0, 6, and 12Hours of exposure to NaCl (J194L\_0H, J194R\_0H, J7L\_0H, J7R\_0H, J194L\_6H, J194R\_6H, J194L\_12H, J194R\_12H, J7L\_6H, J7R\_6H, J7L\_12H, J7R\_12H). For each sample, two biological replicates were taken makes it twenty-four as indicated above.

### **Correlation**

Since the selection of salt-tolerant and sensitive germplasms was made base on germination rate, a study of the relationship between germination rates related traits with seedling stage parameters and qRT-PCR with RNA-Seq results were carried out using R-package software [36].

### **Transcriptome sequencing**

A total amount of 1µg RNA per sample was used as input material for the RNA sample preparations. Twenty-four RNA sequencing libraries were generated using NEB Next Ultra TM RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample [37]. The quality of the library was assessed using the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform, and

paired-end reads were generated for transcriptome sequencing. All sequencing data were deposited into the NCBI database under the Sequence Read Archive (SRA) submission number SUB8495477.

### **Transcriptome analysis and annotation**

Low-quality reads and those containing adapter were removed using in-house Perl scripts. High-quality clean data were used for all subsequent analyses. Transcriptome assembly for cleaned data was performed in Trinity with default parameters [38]. All clean reads were then mapped to the transcripts; those with less than 5X coverage were removed. About eight public databases were used for gene function annotations such as; KO (KeggOrthology), KOG (Eukaryotic Orthologous Groups), Pfam (Homologous protein family), NR (non-redundant database; NCBI), GO (Gene Ontology) and SwissProt (A manually annotated and reviewed protein sequence database) using BLASTx with an E value threshold of  $10^{-5}$  [39], COG (Clusters of Orthologous Groups of proteins) and eggNOG.

### **Biological analysis of differentially expressed genes**

A Gene expression level of each sample was estimated based on RSEM [40]. Clean data reads were then mapped back onto the assembled transcriptome, and thus, a read count per gene was later obtained from the mapping results. DESeq R package (1.10.1) was used to obtain Differentially Expressed Genes (DEGs) of the control and treatment groups [41]. Benjamini and Hochberg's approach was used to adjust the P-value for controlling the false discovery rate. The adjusted P-value  $< 0.01$  found by DESeq in any genes were assigned as differentially expressed. Gene Ontology (GO) enrichment analysis of the DEGs was implemented by the Goseq R packages based on Wallenius non-central hypergeometric distribution [42]. We used KOBAS software to test the statistical enrichment of differentially expressed genes in KEGG pathways.

### **Quantitative reverse transcription-PCR Analysis**

The gene expression from eight randomly selected DEGs was analyzed using a two-step quantitative reverse transcription-PCR (qRT-PCR) to validate the RNA-seq results. Two independent biological and three technical replicates were performed. First, one  $\mu\text{g}$  total RNA per sample was reverse-transcribed into first-strand cDNA using the GoScript<sup>TM</sup> Reverse Transcriptase kit (Promega Corporation 2800 Woods Hollow Road Madison, WI 53711-5399 USA), following manufacturer protocol. After 10 x dilution, cDNA was used as templates for qRT-PCR (Bio-Rad CFX96 Real-Time System C1000 Touch Thermal Cycle USA). The reaction mixture was prepared using the Fast Start Universal SYBR Green Master (ROX) kit (ROCHE) following manufacturer protocol. The jute ELF (elongation factor 1 alpha) gene was selected as the endogenous control [43]. Primers for the DEGs and ELF are listed in Supplementary File (Additional Table S1). Relative expression levels were determined using the comparative Ct method [44].

### **Candidate genes validation**

To validate the salt-tolerant candidate genes, seven genes that were used to validate the RNA-seq above were randomly selected and used in the germplasms (J194 and J7) roots and leaves samples exposed to

NaCl stress at one and two weeks. The salt-tolerant and sensitive germplasms (J194 and J7), were grown on a separate pot filled with Pindrustup media, which were subjected to control and treated (NaCl) conditions. The growing plants were subjected to reasonable conditions and watered; at two leaves stage (two weeks), 50mL of the salt solutions were applied one-day interval to the intended treated pots for another two weeks. At one and two weeks (Additional Figure S7c-d) of the treatment application, leaves, and roots, including the shoots from each sample (J194 and J7) under control and salt-treated conditions, were harvested, frozen immediately in liquid nitrogen, and then stored at -80°C for RNA extraction. Three biological replicates of each sample were used for RNA extraction.

### **Identification of PP2C genes in *Corchorus capsularis* and *Arabidopsis thaliana***

All Arabidopsis PP2C gene families were initially search using the keyword “protein phosphatase 2C” in the NCBI (<http://www.ncbi.nlm.nih.gov>) and TAIR (<http://www.Arabidopsis.org>) databases. The protein sequences obtained were then pooled, and the redundant were removed using a custom Perl program. The Remaining protein sequences were used as a query to perform multiple database searches against genome and proteome files downloaded from TAIR. All redundant sequences were further deleted; the sequences obtained were also verified using the NCBI-Conserved Domain database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) search program and SMART databases (<http://smart.embl-heidelberg.de/>) [45]. Proteins without PP2C Domains’ and those with obvious error on their gene length or less than 100bp were removed. The remaining Arabidopsis protein sequences with full PP2C domains were used as a query. We performed a local blast search against the *Corchorus capsularis* reference genome; those with the highest percentage identity and e-value  $\leq 1e-003$  were selected for further analysis. PP2C Conserved domains were also conducted as stated above, and all redundant sequences were removed accordingly.

### **Alignment and phylogenetic analysis of PP2C sequences**

Amino acid sequences of Arabidopsis and *Corchorus capsularis* PP2C proteins were further investigated, and multiple sequence alignment was conducted by MUSCLE [46] using MEGA7 software with the default options [47]. Maximum likelihood was used while constructing the phylogenetic tree using the full likelihood method; 1000 bootstrap replication values were performed; all other parameters were set as default.

### ***Corchorus capsularis* PP2C Proteins Motifs and genes structure identifications**

*Corchorus capsularis* conserved motif scanning of PP2C proteins was conducted using online tools Multiple Em for Motif Elicitation (MEME) (<http://meme-suite.org/tools/meme>) and [https://salad.dna.affrc.go.jp/CGViewer/en/cgv\\_upload.html](https://salad.dna.affrc.go.jp/CGViewer/en/cgv_upload.html) with default settings, however number of motifs to find was set as 11 and 300 maximum widths. Gene Structure Display Server (GSDS 2.0) (<http://gsds.cbi.pku.edu.cn>) [48] was used for the construction of gene structure.

### **Chromosomal localization and duplication of PP2C in *Corchoruscapsularis***

For the chromosomal location of *Corchorus capsularis* PP2C, an online tool (<http://mg2c.iask.in/mg2c v2.0/>) was used, all parameters were set as default.

### **qRT-PCR analysis of PP2C (reference marker) genes expression in J194 and J7**

Expression analysis of salt stress-responsive PP2C key genes such as; RAB18, RD29B, KIN1, and RD29A were searched and obtained in Arabidopsis TAIR (<http://www.Arabidopsis.org>). The obtained genes were selected according to the e-value and the identity for the next analysis. These genes were then used as query and performed local blast search against *Corchorus capsularis* reference genome from which the following genes according to their identity and e-value was selected for the expression analysis; Cc.04G0022370 (RAB18), Cc.03G0029910 (RD29B), Cc.01G0008270 (KIN1), and Cc.03G0029910 (RD29A). Primers of these genes are presented in (Additional Table S2). The qRT-PCR for the J194 and J7 root and leaf samples (as illustrated above at 0, 6, and 12-hour of exposure to NaCl) were carried out. The samples were also taken for qRT-PCR analysis at one and two weeks of salt treatment in the potting experiment, as stated previously.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable

### **Consent for publication**

Not applicable

### **Availability of data and materials**

The raw transcriptome data were deposited in the NCBI database under the Sequence Read Archive (SRA) submission number of SUB8495477.

### **Competing interests**

The authors have declared that no conflict of interest.

### **Funding**

National Natural Science Foundation of China supported this work (31771369) <http://www.nsf.gov.cn/> as well as Foreign Cooperation Project of Science and Technology Department in Fujian, China (2015I0001) <http://xmgl.fjkjt.gov.cn/>.

### **Authors' contributions**

AKI: Investigation, Writing, Formal analysis, Editing. YX: Resources, Data curation, Formal Analysis. QH: Editing. SN: Editing. MZA: Editing. LZ: Editing. JQ: Supervision. LZ: Software, Validation, Supervision,

Funding acquisition, and edited the manuscript.

## Acknowledgments

We acknowledged the assistance of all members of the Liwu Zhang lab.

## References

1. Ma, H., et al., *Differential proteomic analysis of salt stress response in jute (Corchorus capsularis & olitorius L.) seedling roots*. Pak J Bot, 2015. **47**(2): p. 385-96.
2. Ogunkunle, C.O., et al., *Bioaccumulation and associated dietary risks of Pb, Cd, and Zn in amaranth (Amaranthus cruentus) and jute mallow (Corchorus olitorius) grown on soil irrigated using polluted water from Asa River, Nigeria*. Environmental monitoring and assessment, 2015. **187**(5): p. 281.
3. Zhu, J.-K., *Abiotic stress signaling and responses in plants*. Cell, 2016. **167**(2): p. 313-324.
4. Li, S., et al., *Effects of drought and salt-stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-seq*. BMC genomics, 2016. **17**(1): p. 200.
5. Munns, R., *Genes and salt tolerance: bringing them together*. New phytologist, 2005. **167**(3): p. 645-663.
6. Cheng, C., et al., *SCFAtPP2-B11 modulates ABA signaling by facilitating SnRK2.3 degradation in Arabidopsis thaliana*. PLoS genetics, 2017. **13**(8): p. e1006947.
7. de Zelicourt, A., J. Colcombet, and H. Hirt, *The role of MAPK modules and ABA during abiotic stress signaling*. Trends in plant science, 2016. **21**(8): p. 677-685.
8. Fujii, H. and J.-K. Zhu, *Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress*. Proceedings of the National Academy of Sciences, 2009. **106**(20): p. 8380-8385.
9. Furihata, T., et al., *Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1*. Proceedings of the National Academy of Sciences, 2006. **103**(6): p. 1988-1993.
10. Miller, G., et al., *Reactive oxygen species homeostasis and signalling during drought and salinity stresses*. Plant, cell & environment, 2010. **33**(4): p. 453-467.
11. Li, D.-M., et al., *Ferulic acid pretreatment enhances dehydration-stress tolerance of cucumber seedlings*. Biologia plantarum, 2013. **57**(4): p. 711-717.
12. Sofo, A., et al., *Ascorbate peroxidase and catalase activities and their genetic regulation in plants subjected to drought and salinity stresses*. International Journal of Molecular Sciences, 2015. **16**(6): p. 13561-13578.
13. Peng, Z., et al., *Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes*. BMC genomics, 2014. **15**(1): p. 760.
14. Cao, J., et al., *Genome-wide identification and evolutionary analyses of the PP2C gene family with their expression profiling in response to multiple stresses in Brachypodium distachyon*. BMC genomics,

2016. **17**(1): p. 175.
15. Haider, M.S., et al., *Insights into grapevine defense response against drought as revealed by biochemical, physiological and RNA-Seq analysis*. Scientific reports, 2017. **7**(1): p. 1-15.
  16. Murata, Y., et al., *Abscisic acid activation of plasma membrane Ca<sup>2+</sup> channels in guard cells requires cytosolic NAD (P) H and is differentially disrupted upstream and downstream of reactive oxygen species production in abi1-1 and abi2-1 protein phosphatase 2C mutants*. The Plant Cell, 2001. **13**(11): p. 2513-2523.
  17. Kerk, D., et al., *The complement of protein phosphatase catalytic subunits encoded in the genome of Arabidopsis*. Plant physiology, 2002. **129**(2): p. 908-925.
  18. Singh, A., et al., *Protein phosphatase complement in rice: genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development*. BMC genomics, 2010. **11**(1): p. 1-18.
  19. Nakashima, K. and K. Yamaguchi-Shinozaki, *ABA signaling in stress-response and seed development*. Plant cell reports, 2013. **32**(7): p. 959-970.
  20. Yang, R., et al., *Hormone profiling and transcription analysis reveal a major role of ABA in tomato salt tolerance*. Plant physiology and biochemistry, 2014. **77**: p. 23-34.
  21. Xue, T., et al., *Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis*. BMC genomics, 2008. **9**(1): p. 550.
  22. Fracasso, A., L.M. Trindade, and S. Amaducci, *Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE*. BMC Plant Biology, 2016. **16**(1): p. 115.
  23. Lei, L., et al., *Identification of a Major QTL and Candidate Gene Analysis of Salt Tolerance at the Bud Burst Stage in Rice (Oryza sativa L.) Using QTL-Seq and RNA-Seq*. Rice, 2020. **13**(1): p. 1-14.
  24. Yang, Z., et al., *Salt-Stress response mechanisms using de Novo transcriptome sequencing of salt-tolerant and sensitive Corchorus spp. genotypes*. Genes, 2017. **8**(9): p. 226.
  25. Finkelstein, R.R., S.S. Gampala, and C.D. Rock, *Abscisic acid signaling in seeds and seedlings*. The Plant Cell, 2002. **14**(suppl 1): p. S15-S45.
  26. Wang, S., et al., *Auxin-related gene families in abiotic stress response in Sorghum bicolor*. Functional & integrative genomics, 2010. **10**(4): p. 533-546.
  27. Swarup, R., et al., *Structure-function analysis of the presumptive Arabidopsis auxin permease AUX1*. The Plant Cell, 2004. **16**(11): p. 3069-3083.
  28. Zhang, F., et al., *Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, Gossypium davidsonii*. Scientific reports, 2016. **6**: p. 20582.
  29. Kuhn, J.M., et al., *The protein phosphatase AtPP2CA negatively regulates abscisic acid signal transduction in Arabidopsis, and effects of abh1 on AtPP2CA mRNA*. Plant physiology, 2006. **140**(1): p. 127-139.
  30. Liu, P.F., et al., *Signaling pathways mediating the suppression of Arabidopsis thaliana Ku gene expression by abscisic acid*. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2008. **1779**(3): p. 164-174.

31. Singh, A., et al., *ABA inducible rice protein phosphatase 2C confers ABA insensitivity and abiotic stress tolerance in Arabidopsis*. PloS one, 2015. **10**(4): p. e0125168.
32. Shazadee, H., et al., *Identification and Expression Profiling of Protein Phosphatases (PP2C) Gene Family in Gossypium hirsutum L.* International journal of molecular sciences, 2019. **20**(6): p. 1395.
33. Cannon, S.B., et al., *The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana*. BMC plant biology, 2004. **4**(1): p. 10.
34. Cheung, J., et al., *Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence*. Genome biology, 2003. **4**(4): p. R25.
35. Prince, V.E. and F.B. Pickett, *Splitting pairs: the diverging fates of duplicated genes*. Nature Reviews Genetics, 2002. **3**(11): p. 827-837.
36. Wiley, M. and J.F. Wiley, *Installing R*, in *Beginning R 4*. 2020, Springer. p. 1-19.
37. Yang, Z., et al., *De novo transcriptome sequencing of two cultivated jute species under salinity stress*. PloS one, 2017. **12**(10): p. e0185863.
38. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. Nature biotechnology, 2011. **29**(7): p. 644.
39. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic acids research, 1997. **25**(17): p. 3389-3402.
40. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*. BMC bioinformatics, 2011. **12**(1): p. 323.
41. Wang, L., et al., *DEGseq: an R package for identifying differentially expressed genes from RNA-seq data*. Bioinformatics, 2010. **26**(1): p. 136-138.
42. Young, M.D., et al., *Gene ontology analysis for RNA-seq: accounting for selection bias*. Genome biology, 2010. **11**(2): p. R14.
43. Ferdous, A.S., et al., *Identification of stable reference genes for quantitative PCR in jute under different experimental conditions: An essential assessment for gene expression analysis*. Australian Journal of Crop Science, 2015. **9**(7): p. 646.
44. Livak, K. and T. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$  method*. *methods*. 2001; **25**: 402-408. External Resources Pubmed/Medline (NLM) Crossref (DOI), 2001.
45. Letunic, I., T. Doerks, and P. Bork, *SMART 7: recent updates to the protein domain annotation resource*. Nucleic acids research, 2012. **40**(D1): p. D302-D305.
46. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucleic acids research, 2004. **32**(5): p. 1792-1797.
47. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets*. Molecular biology and evolution, 2016. **33**(7): p. 1870-1874.
48. Hu, B., et al., *GSDS 2.0: an upgraded gene feature visualization server*. Bioinformatics, 2015. **31**(8): p. 1296-1297.