De novo genome assembly and *in natura* epigenomics reveal salinity-induced DNA methylation in the mangrove tree *Bruguiera gymnorhiza*

Matin Miryeganeh[1] (iD), Ferdinand Marlétaz[2] (iD), Daria Gavriouchkina[3] (iD) and Hidetoshi Saze[1] (iD)

[1]Plant Epigenetics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan; [2]Department of Genetics, Evolution and Environment (GEE), University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK; [3]Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan

Summary

• Mangroves are adapted to harsh environments, such as high ultraviolet (UV) light, low nutrition, and fluctuating salinity in coastal zones. However, little is known about the transcriptomic and epigenomic basis of the resilience of mangroves due to limited available genome resources.

• We performed a *de novo* genome assembly and *in natura* epigenome analyses of the mangrove *Bruguiera gymnorhiza*, one of the dominant mangrove species. We also performed the first genome-guided transcriptome assembly for mangrove species.

• The 309 Mb of the genome is predicted to encode 34 403 genes and has a repeat content of 48%. Depending on its growing environment, the natural *B. gymnorhiza* population showed drastic morphological changes associated with expression changes in thousands of genes. Moreover, high-salinity environments induced genome-wide DNA hypermethylation of transposable elements (TEs) in the *B. gymnorhiza*. DNA hypermethylation was concurrent with the transcriptional regulation of chromatin modifier genes, suggesting robust epigenome regulation of TEs in the *B. gymnorhiza* genome under high-salinity environments.

• The genome and epigenome data in this study provide novel insights into the epigenome regulation of mangroves and a better understanding of the adaptation of plants to fluctuating, harsh natural environments.

Introduction

Mangrove trees have evolved special characteristics, such as vivipary and aerial roots, to adapt to harsh intertidal environments and are resilient to many environmental stresses, including salinity, intense ultraviolet (UV) radiation, and anaerobic soils (Tomlinson, 1986). Their high adaptability makes them a valuable model for understanding the molecular mechanisms underlying stress tolerance and adaptation in plants (Dassanayake *et al.*, 2010; Xu *et al.*, 2017). Previous studies have identified various genes and metabolic pathways in stress responses that are implicated in the adaptive evolution of mangroves to the extreme environments (Miyama & Hanagata, 2007; Dassanayake *et al.*, 2009; Chen *et al.*, 2011; Yang *et al.*, 2015; Guo *et al.*, 2017; Feng *et al.*, 2020; Meera & Augustine, 2020). However, mangrove trees often show drastic changes in phenotypes, such as biomass and tree height, in response to a spectrum of stress factors, including high salinity, strong wind, and poor nutrient conditions (Suwa *et al.*, 2009), suggesting the involvement of epigenome regulation in their phenotypic plasticity in natural environments.

Among various epigenome modifications, DNA cytosine methylation has been intensively studied in plants, which primarily targets and silences repeats and transposable elements (TEs) in plant genomes (Slotkin & Martienssen, 2007). DNA methylation is regulated by multiple chromatin modifications (Law & Jacobsen, 2010; Matzke & Mosher, 2014). Small interfering RNAs (siRNAs) generated by RNA interference (RNAi) machinery and *de novo* methylase DOMAIN REARRAENGED METHYLTRANSFERASEs (DRMs) are involved in the RNA-directed DNA methylation (RdDM) pathway that establishes DNA cytosine methylation at CG and non-CG (CHG and CHH contexts; H can be A, or C, or T) sites. However, histone H3 Lysine9 methylation (H3K9me) mediated by histone methyltransferases recruits CHROMOMETHYLASEs (CMTs) to introduce non-CG methylation at TEs. In contrast, active DNA demethylation is mediated by DNA glycosylases that remove TE methylation, especially in the reproductive tissues of plants (Zhang *et al.*, 2018). DNA methylation changes are often induced in response to biotic and abiotic stresses, which are known to modulate gene expression and phenotypic variations (Dowen *et al.*, 2012; Sahu *et al.*, 2013; Soubry, 2015; Lamke & Baurle, 2017), potentially leading to stress resistance and adaptation in plants (Wibowo *et al.*, 2016; Erdmann & Picard, 2020).

Despite the potential importance of epigenome regulation in the environmental adaptations of mangroves, there has only been a few epigenome studies on mangroves (Lira-Medeiros *et al.*, 2010; Wang *et al.*, 2018), mainly due to the limited genome information on diverse mangrove species (Xu *et al.*, 2017; Hu *et al.*, 2020). To investigate the role of epigenome regulation in stress responses and adaptations of these trees, especially in natural environments, we assembled the genome and analyzed the transcriptome and epigenome of natural populations of *Bruguiera gymnorhiza* (L.) Lamk., which is one of the most widely distributed mangrove trees from the most mangrove-rich family Rhizophoraceae (Duke & Ge, 2011; Christenhusz & Byng, 2016). Same as all the species belonging to the family Rhizophoracee, *B. gymnorhiza* is diploid and possesses 36 chromosomes ($2n = 36$). The flowers are bisexual, self-compatible, and self/cross-pollinating (Raju Aluri, 2014). The *de novo* assembled *B. gymnorhiza* genome is approximately 309 Mb in size, encoding 34 403 genes and consisting of 48% of repetitive sequences. Intriguingly, the plants growing under high-salinity conditions showed DNA hypermethylation in TEs in both natural and controlled experiments, suggesting suppression of TE activation under the stress conditions. Our study provides novel insights into the epigenome regulation of stress responses of mangroves that are highly resilient to extreme natural environments.

Materials and Methods

De novo genome assembly

Plant material, and genomic DNA extraction method is given in the Supporting Information Notes S1.

Estimation of the genome size

Flow cytometry Leaves (20–40 mg) were chopped in 1–2 ml ice-cold nuclear isolation buffer (45 mM $MgCl_2$, 20 mM MOPS, 30 mM sodium citrate, 0.01% (v/v) Triton X-100, 5 mM beta mercaptoethanol/1 ml), with a disposable scalpel. The homogenate was mixed by pipetting several times, and the debris was filtered into a sample tube through a 42 μm nylon mesh. Finally, propidium iodide (DNA fluorochrome; Thermo Fisher Scientific, Waltham, MA, USA) (50 μg ml^{-1}) and RNase A (Thermo Fisher Scientific) were added. Stained nuclei were analyzed using a FACSCalibur flow cytometer. Values of nuclear DNA were estimated by comparing the nuclear peak on the linear scale with the peak for *Arabidopsis thaliana* and *Solanum lycopersicum* included as internal standards (Fig. S1a).

k-mer analysis A total of 72 Gb of reads obtained from the PE Illumina Hi-seq platforms were subjected to a 27-mer frequency distribution analysis with JELLYFISH v.2.2.3 (Marcais & Kingsford, 2011). The analysis parameters were set as -t 8 -C -m 27 -s 5G, and the final result was plotted as a frequency curve (Fig. S1b). One peak was observed from the distribution curve, which provided a peak depth of 51 for the estimation of its genome size. Since the total number of k-mers was

18 636 738 543, the genome size was calculated to be approximately 373 Mb using the formula: genome size = k-mer number/peak depth. This indicated that the sequenced short Illumina reads provided *c.* ×51 coverage. The genome size and homogenity was then investigated further by running GENOMESCOPE 2.0 (Ranallo-Benavidez *et al.*, 2020). The result supports a diploid genome with a proportion of heterozygosity carried by paralogs (i.e. non-diploid k-mer pairs) of 0.37 (Fig. S1b).

Whole-genome sequencing

Genome sequencing was performed using a combination of PacBio RS II and Illumina sequencers.

Details about Illumina sequencing is given as Notes S1.

PacBio sequencing

For PacBio sequencing, high-molecular weight DNA at a concentration of 70 ng μl^{-1} and a sample volume of 350 μl were sent to the SQC-OIST for sequencing. Library preparation for PacBio sequencing was performed with high-quality genomic DNA, and sequencing was performed in four single molecule real-time (SMRT) cells using P6-C4 chemistry (20 kb protocol), resulting in 7651 821 reads, with an N50 read length of 8393 bp and a mean read length of 6524 bp. In total, approximately 50 Gb reads were generated on the PacBio platforms, representing *c.* ×120 coverage of the genome. Altogether, the three libraries provided 84 Gb of trimmed and filtered sequencing data, which translated to *c.* ×240 coverage of the genome (Tables S1, S2).

Whole-genome assembly

For *de novo* assembly of the reads, with deep coverage from each sequencing platform, we initially used three strategies: (1) Illumina reads alone; (2) PacBio reads alone; and (3) combination of Illumina and PacBio reads. Thereafter, we compared the genome statistics for each strategy (Table S3) and chose the combination of Illumina and PacBio for assembly as the best strategy, as outlined later.

Assembly of PE and MP Illumina data is given as Notes S1

Assembly of PacBio single-molecule long reads

Additionally, another assembly was built using only PacBio long reads with the program CANU-1.7.1 (Koren *et al.*, 2018), resulting in 4008 contigs with an N50 length of 1.3 Mb. Genome polishing was performed using the Arrow algorithm of the variantCaller tool within the GENOMICCONSENSUS package v.2.3.2 (PacificBiosciences, Menlo Park, CA, USA) using default parameters, to improve site-specific consensus accuracy.

Final assembly using both Illumina and PacBio platform data

Finally, in an effort to construct the best assembly, all MP and PE Illumina reads were used to scaffold the resulting contigs from

PacBio, assembly only by SSPACE/sspace_basic-2.1.1 (Boetzer *et al.*, 2011), yielding 2282 scaffolds and an N50 size of 1.5 Mb. Finally, the size of the scaffold N50 was extended to 2.3 Mb, and the number of scaffolds decreased to 1075 using SOAPDENOVO GAPCLOSER (Table S3). Using this assembly pipeline, we obtained a final draft genome of 309 Mb. The quality of the genome assemblies was assessed using QUAST (Gurevich *et al.*, 2013). The final assembly results showed that contig N50 was significantly improved, from 470 16 kb in the Illumina to 2.3 Mb in the final assembly (Table S3).

Evaluation of the assembled genome

Validation of assembled genome was assessed using two different approaches:

(1) Quality trimmed Illumina HiSeq PE reads were aligned to our assembly using BWAMEM (Li & Durbin, 2009) and the mapping efficiency was checked using ALFRED (Thankachan *et al.*, 2016). Moreover, single nucleotide polymorphisms (SNPs) and small indels were called by FREEBAYES (Garrison & Marth, 2012). The matching rate was > 99% (Table S4) and a unimodal coverage dataset was observed which corroborate the homogeneity of the assembly. The variant call formatted (VCF) SNP file was then analyzed using BCFTOOLS (https://github.com/samtools/bcftools) to select SNPs in homozygous positions with the minimum quality score of 20 and the coverage cut-off of 20. This identified only 22 984 SNPs, which means only 0.007% of the positions in the genome were affected and gave us confidence that no further polishing was necessary. (2) We compared the genome assembly against a set of Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simao *et al.*, 2015) to test the completeness of the genome assembly and gene space using the plant-specific profile. This approach makes use of the single-copy genes expected to be present in plants (1440 total BUSCO groups were searched). Consequently, 94.2% (1354) of the gene set was 'completely' retrieved (Table S5). To assess the completeness level of gene annotations, we also compared the predicted proteins of the assembled genome against BUSCO in protein mode with the eudicots_odb10 data set. The result showed that 93.5% (2133/2326) of the potential proteins were 'completely' retrieved which shows the good quality of the annotation of the genome (Table S5). Only 3.8% missing genes shows that the annotation has captured almost all the coding genes in *B. gymnorhiza* genome (Table S5).

Genome annotation

REPEATMODELER v.2.1 (Flynn *et al.*, 2020) was used with default parameters for *de novo* repeat identification in the assembled genome (Table S6); the constructed raw repeat library was used as the input for REPEATMASKER (v.4.1.0, https://www.repeatmasker.org) to classify the types of repetitive sequences and repeat sequences were masked throughout the genome before further analysis. The LTR-FINDER software (Xu & Wang, 2007) was also used to specifically search for full-length LTR retrotransposons in the genome. The results were then compared to the REPEATMODELER

output (Table S7), and as there were no new LTR found by the LTR-Finder, we considered the REPEATMODELER output a good source for our reference in this study.

We used the combination of natural and control samples (all 25 libraries) for transcriptome assembly. Reads from 25 PE RNA-sequencing (RNA-Seq) were mapped on the repeat masked genome using STAR (Dobin *et al.*, 2013), and the mapping results were then assembled into potential transcripts using STRINGTIE v.1.3.6 (https://ccb.jhu.edu/software/stringtie/). The assembly of clean reads produced the total of 131 108 transcripts with an average length of 1400 bp and N50 of 2150 bp (Table S3). The assembled transcripts from each sample were then merged using TACO (Niknafs *et al.*, 2017), and the resulting transcriptome was subjected to prediction of coding DNA sequences (CDSs) using TRANSDECODER- v.5.5.0 (http://transdecoder.sf.net) built in scripts ('gtf_genome_to_cdna_fasta.pl' and 'gtf_to_alignment_gff3.pl'). We then used PORTCULLIS-v.1.1.2 (Mapleson *et al.*, 2018) and 'bam2hints' built in script from AUGUSTUS-v.3.3 (http://bioinf.uni-greifswald.de/augustus) on mapped RNA bam files and the genome to get intron hits. Additionally, we used an in-house Python script on the resulting transcriptome from TACO to obtain exon hints (Notes S2). The intron and exon hints were then used to train AUGUSTUS. Coding sequences were obtained using CUFFLINKS 'gffread' utility (http://cole-trapnell-lab.github.io/cufflinks/file_formats/#the-gffread-utility) and were aligned against the National Center for Biotechnology Information (NCBI) nonredundant protein databases using BLAST (Wheeler *et al.*, 2007), with an e-value threshold of 1×10^{-6} to annotate the functions of genes. Finally, the genome was found to contain 34 403 genes, with 10 140 identified as being similar to known genes (Table S8).

Gene ontology (GO) annotation was obtained by aligning against the UNIPROT database using DAVID 6.8 (https://david.ncifcrf.gov/home.jsp), where the protein sequences were also searched against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database for KEGG orthology (KO) assignments and pathway annotation. The output data of all GO annotations were thoroughly examined to identify those involved in stress, growth, and epigenetics. The *B. gymnorhiza* chloroplast genome was assembled using GETORGANELLE v.1.7.1 (Jin *et al.*, 2020) and visualized by GESEQ (Tillich *et al.*, 2017). Graphs for GO biological process, cellular component, and molecular function were plotted using Hayai-Annotation Plants (Ghelfi *et al.*, 2019). Transfer RNAs (tRNAs) were predicted by tRNAScan-SE 2.0.6 (Chan & Lowe, 2019) with default settings. Ribosomal RNA (rRNA), microRNA (miRNA), and small nucleolar RNA (snoRNA) were predicted by INFERNAL 1.1.2 (Nawrocki & Eddy, 2013) against the Rfam.cm database. The *B. gymnorhiza* 14 842 EST dataset (Miyama *et al.*, 2006) (DDBJ accession numbers BP938635 to BP953476) was used for homology search using BLASTN with default settings by CLC Genomics Workbench 20.0.4 (Qiagen). The genome data of *Kandelia obovata* (Hu *et al.*, 2020) and *Rhizophora apiculata* (Xu *et al.*, 2017) were downloaded from the databases under accession numbers GWHACBH00000000 and PRJEB8423, respectively. A phylogenetic tree was generated by MEGAX (Kumar *et al.*, 2018) using CDSs of the plastid genomes of plant species. DNA sequences

were aligned using Muscle with default settings, and a phylogenetic tree was constructed using the maximum likelihood method with 1000 bootstrap reiteration, -nucleotide, and Kimura 2-parameter model. Synteny blocks of Mangrove genomes (scaffolds/contigs longer than 5 Mb) were identified using SyMap v.5.0.6 (Soderlund et al., 2006) with default settings (Table S9), and a synteny map was plotted in RIdeogram (Hao et al., 2020). Syntenic regions of the BURP gene cluster between B. gymnorhiza and K. obovata genomes were searched by generating genome files in the GenBank format using the seqret function in EMBOSS (Rice et al., 2000), and visualized using Easyfig 2.2.2 (Sullivan et al., 2011) with the following settings: Blastn; minimum length = 300; maximum e-value = $1e^{-5}$; minimum identity value = 80.

Plant material for natural and control experiments

Samples for 'natural' experiments were obtained from the same mangrove forest where the individual for genome assembly was collected from. In the natural study site along the Okukubi River, Okinawa Island, Japan (26°27′N, 127°56′E), the riverside trees were located along the estuary of the river, and the oceanside trees grew along the coastal area of the Japanese Pacific Ocean. Propagules of the individual B. gymnorhiza tree in the natural environment that was used for the whole genome assembly, were all collected at the same day, and were used for all the individual plants in control experiments. The control experiment was performed in the laboratory, where samples grown in saline water were compared to samples grown in brackish water (hereafter referred to as saline and brackish, respectively) in the laboratory inside a growth chamber. We planted B. gymnorhiza seedlings in 25 cm diameter plastic pots and placed them in stainless steel tanks filled with brackish (1.5% salinity) or saline (4% salinity) water to simulate the 'riverside' and 'oceanside' natural environments.

Whole-genome bisulfite sequencing and DNA methylation analysis

Leaves from two biological individual replicates from each condition (two of each oceanside and riverside for natural experiment and two of each saline and brackish for the control experiment; in total, eight samples) were collected, frozen in liquid nitrogen, and stored at −80°C for DNA extraction. Intact genomic DNA extracted from leaves (using the same extraction method as described earlier) was used for whole-genome bisulfite sequencing (WGBS) with the post-bisulfite adapter tagging (PBAT) method (Miura et al., 2012) on the Illumina Hi-Seq 2500 platform (Illumina, San Diego, CA, USA) at OIST-SQC. After removing low-quality reads using Trimmomatic-0.36 (Bolger et al., 2014), the quality-trimmed reads were aligned to the de novo built genome using Bismark v.0.19.0 (Krueger & Andrews, 2011) with default parameters. Only uniquely mapped reads were retained, and only cytosines covered by at least three reads were retained and the two replicates were then intersected using the bedtools intersect option (BEDTools v.2.27.1, https://github.com/arq5x/bedtools2) for further analysis. After removing duplicate reads, the mapped reads were merged to calculate the counts of each potentially methylated cytosine. The conversion rate (rate at which unmethylated cytosines were converted to uracil by bisulfite) was calculated and cytosines were termed methylated (false discovery rate, FDR < 0.01) using a binomial test employing the conversion rate at the scaffold420 with high homology to chloroplast genomes in other species, followed by Benjamini–Hochberg (Benjamini & Hochberg, 1995) multiple testing correction.

Differentially methylated regions (DMRs) in CG, CHG, and CHH contexts were detected using CGmapTools (Guo et al., 2018) (cgmaptools intersect) and (cgmaptools dmr) with the following parameters: -c 3 -s 200 -S 4000. Detected DMRs (FDR < 0.01) were further filtered by ≥|0.4| change in CG, ≥|0.2| change in CHG, and ≥|0.1| change in CHH contexts. For TE analysis, the same TE family within 50 bp was merged, and copies less than 50 bp in length were excluded from the analysis. DNA methylation distribution plots were performed with deepTools (Ramirez et al., 2016). Venn diagrams were drawn using the eulerr web tool (http://eulerr.co). Heatmaps were generated using Pheatmap1.0.12 (https://cran.r-project.org/web/packages/pheatmap/) with default parameters. The Circos plot was plotted using the Clico FS (Cheong et al., 2015). Boxplots and violin plots were plotted using GPUBR (https://CRAN.R-project.org/package=ggpubr).

Genome features associated with DMRs were identified by bedtools intersect with -f 0.5 parameter in the following order: TE > Gene > Intergenic. DMRs associated with genes and flanking regions were extracted by bedtools intersect the -f 0.9 setting in the following order: Genebody > 0–1 kb upstream of Genebody > 0–1 kb downstream of Genebody > 1–3 kb upstream of Genebody > 1–3 kb downstream of Genebody. Genes overlapping > 80% with TE annotation were excluded from the analysis.

Details about RNA extraction and RNA-Seq is given at Notes S1.

Transcription analyses

The quality treamed RNA-Seq reads were aligned to the draft genome using STAR v.2.7 (Dobin et al., 2013) with a 1-bp mismatch (Table S10), and counts of reads uniquely mapped to annotated genes were obtained using FeatureCounts (v.1.6.3) (Liao et al., 2014). Briefly, the measurement of gene expression level for each transcript was performed using the fragments per kilo bases or per million bases of transcripts, using the mapped transcripts per million (TPM) (Gongora-Castillo et al., 2012). Differential expression analysis was performed using DESeq2 in R software (v.3.6.1) (Love et al., 2014). FDR values ≤ 0.05 and |$\log_2$ FC| ≥ 1 were considered the threshold of differential expression (Zhao et al., 2018). We used the David Gene Ontology (https://david.ncifcrf.gov/home.jsp) for functional analysis of expressed homologous gene pairs to determine overrepresented GO categories across biological processes, cellular components, and molecular function domains. Enrichment of GO terms were tested using Fisher's exact test, with $P < 0.05$ considered as significant. Statistical analyses and visualizations were performed using R. Venn diagrams were drawn using the Eulerr web tool (http://

eulerr.co), and heatmaps with clustering were generated using PHEATMAP1.0.12, with the default parameters (https://cran.r-project.org/web/packages/pheatmap/). Volcano plots were generated using ENHANCEDVOLCANO (https://github.com/kevinblighe/EnhancedVolcano). RNA-Seq data were converted to bigwig files using DEEPTOOLS (Ramirez *et al.*, 2016) bamCoverage function with options; bs = 1, -normalizeUsing CPM. Data tracks were visualized using the INTEGRATED GENOME BROWSER (IGB v.9.1.6) (Freese *et al.*, 2016). Boxplots and violin plots were plotted using GGPUBR as described earlier.

Results

De novo genome assembly of the mangrove *B. gymnorhiza*

To obtain a reference for the genome, transcriptome, and epigenome analyses of mangrove species, we performed a *de novo* genome assembly of the genome of *B. gymnorhiza* (Fig. 1a), one of the dominant mangrove species with a wide distribution range in the Indo-West Pacific region (Duke & Ge, 2011). Flow cytometry analyses of the nuclei and *k*-mer spectrum ($k = 27$) analysis of the Illumina short-read sequences indicated that the genome size of *B. gymnorhiza* was estimated to be *c.* 365 Mb (Fig. S1a,b). Combining PacBio long reads (total 50 Gb), Illumina paired-end (PE) reads (56 Gb), and Illumina mate-pair reads (44 Gb) (Tables S1–S3), we obtained a final assembly containing 1075 scaffolds with a total size of 309 Mb genome (Table S3). The smaller size of the assembled genome compared to the predicted size from *k*-mer and flow cytometry analyses

could be either because of collapsing similar regions in scaffolds or because of the differences between Illumina and PacBio datasets in handling the repeat regions, as Illumina reads were used for *k*-mer analysis. The longest scaffold was 14.1 Mb, the N50 of contigs was 2.3 Mb, and the 31 and 104 largest scaffolds covered 50% and 75% of the genome, respectively (Table S3). Gene model prediction and annotation based on a repeat-masked genome yielded a total of 34 403 protein-coding genes, of which 10 138 genes showed homology to genes in other plant species (Tables S8, S11; Figs S1c, S2). In addition, tRNA genes (Table S12 and other small and non-coding RNA genes, including miRNAs, snoRNAs and rRNA genes were annotated (Table S13). To further assess the completeness of the assembly, we investigated the recovery of the 1440 plant-conserved gene set of BUSCO, which showed that 94.0% (1354) of the gene set was 'completely' retrieved, and 1.7% (24) was 'partially' retrieved, while 4.3% (62) was 'missing' (Table S5). In addition, 13 085 out of 14 842 previously reported expressed sequence tags (ESTs) for *B. gymnorhiza* (Miyama *et al.*, 2006) matched to the 5138 out of the 34 403 annotated gene models in this study (BLASTN, *e*-value < $1e^{-5}$, Table S8), indicating the identification of thousands of novel genes, in addition to the reported *B. gymnorhiza* ESTs. The assembled genome contained 48.66% repetitive sequences, with 25.46% of the total assembly consisting of common TE families and 23.20% of unclassified repeats (Table S14). Among the TEs, LTR-type retrotransposon families, including Gypsy (16.4%), Copia (15.4%), and Caulimovirus (5.1%) were the most abundant in the genome (Fig. S1d; Tables S14, S15. The *B. gymnorhiza* genome is larger than the recently reported
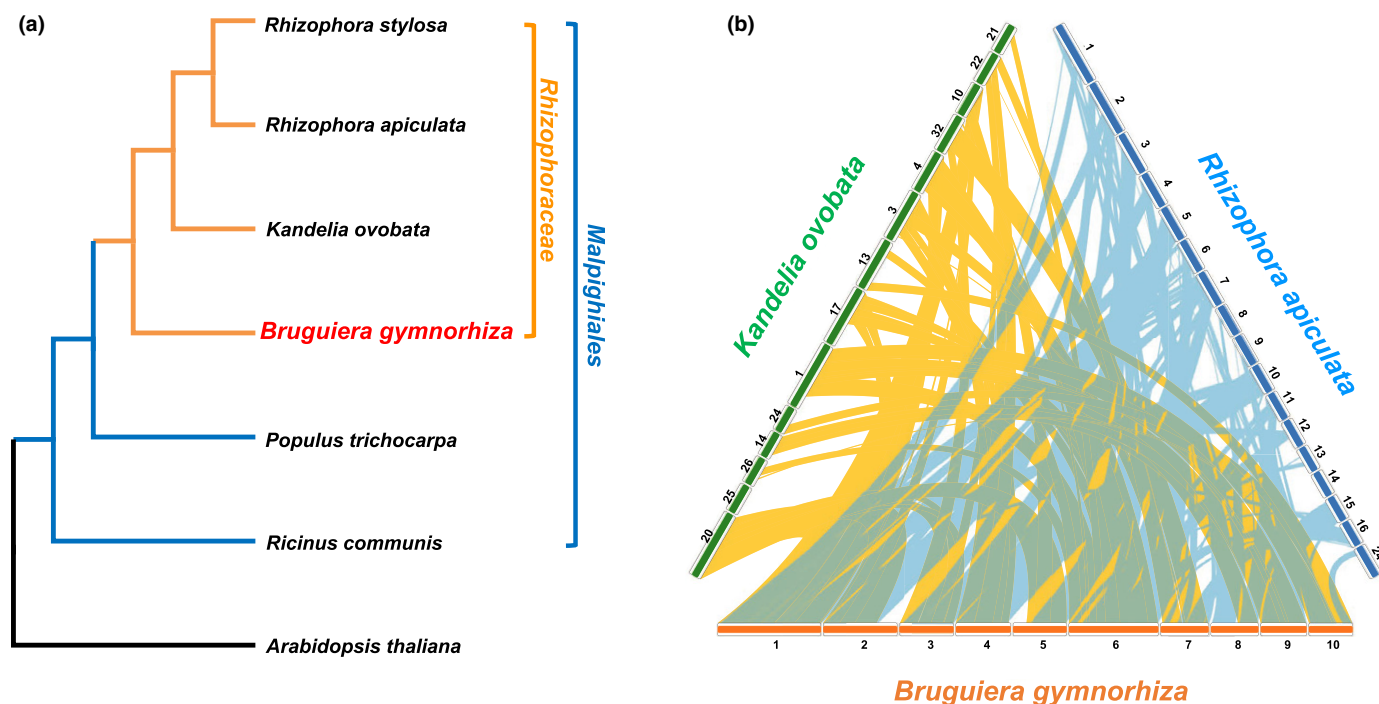

Fig. 1 (a) A cladogram of plant species including *Bruguiera gymnorhiza*. (b) Synteny blocks between *B. gymnorhiza* and *Kandelia obovata* (yellow ribbons), and *B. gymnorhiza* and *Rhizophora apiculata* (light blue ribbons) genomes. Scaffolds/contigs longer than 5 Mb were used for the analysis. Scaffold/contig numbers are indicated.

genome sequences of other mangrove species from the same family Rhizophoraceae, *R. apiculata* (274 Mb) (Xu *et al.*, 2017) and *K. obovata* (178 Mb) (Hu *et al.*, 2020). This might be due to the higher accumulation of repetitive elements in the *B. gymnorhiza* genome. The genome sequence comparisons between *B. gymnorhiza* and *R. apiculata*, and *B. gymnorhiza* and *K. obovata* identified conserved synteny blocks between the genomes (Fig. 1b; Table S9). However, the syntenic genomic regions in the *B. gymnorhiza* genome were widely scattered to different scaffolds in the other two species, suggesting extensive rearrangement of mangrove genomes. This is interesting and needs to be investigated further as it might be something specific to the Rhizophoraceae family as they have diverged from their most recent common ancestor only around 40.7 Myr ago (Xu *et al.*, 2017). Since the *B. gymnorhiza* genome in this study is still at a draft state, a pseudochromosome level assembly in the subsequent study by Hi-C (Belton *et al.*, 2012) that can span very distant DNA regions (Peona *et al.*, 2021) would provide us a clearer picture of genome divergence among the mangrove species. Nonetheless, a nearly two-fold genome-size difference between *B. gymnorhiza* (*c.* 365 Mb) and *K. ovobata* (*c.* 170 Mb) (Hu *et al.*, 2020) suggests that there might have been extensive genome rearrangements and expansions/contractions after the divergence of the species.

Differential transcriptome profiles of *B. gymnorhiza* in natural and controlled environments

To investigate the genome-wide gene expression dynamics of *B. gymnorhiza* in natural environments, we performed RNA-Seq analysis using *B. gymnorhiza* samples obtained *in natura*. The experiment was conducted using leaf samples from natural *B. gymnorhiza* populations in Okinawa, Japan (Fig. 2a), where trees were growing in a wide range of salinity conditions, from the oceanside with high salinity to the upper riverside with brackish environments (saline and brackish, hereafter). Trees in the oceanside (saline) generally showed a dwarf phenotype with smaller leaves, while trees growing *c.* 1 km away on the riverside (brackish) were taller, with thicker trunks and larger leaves (Fig. 2b,c), indicating drastic alterations in the morphological phenotypes of *B. gymnorhiza* depending on their growing environments. To corroborate the *in natura* transcriptome analysis, we further conducted RNA-Seq for *B. gymnorhiza* plants grown in a controlled environment in a laboratory growth chamber, where plants were grown under 1.5% or 4% salinity, to simulate the riverside 'brackish' or the oceanside 'saline' natural environments. In total, 25 samples (six biological replicates from the oceanside (saline), five from the riverside (brackish) for natural samples, eight biological replicates from saline, six from brackish conditions for control samples, detailed in Table S16) were sequenced and analyzed. Importantly, the whole-genome transcriptome profiles of both natural and control saline samples, or natural and control brackish samples, showed high similarities (Figs 2d, S3a), suggesting that the salinity difference in the control conditions largely mimicked the natural conditions that resulted in the transcriptome difference. We identified 5467 upregulated (> two-fold, *P*-adjusted < 0.05, Table S17) and 2242 downregulated genes (< 0.5-fold, *P*-adjusted < 0.05, Table S18) in the natural saline environment compared to the brackish environment (Fig. 2e), including previously reported salt stress-responsive genes of *B. gymnorhiza* such as Bg70 (Miyama & Hanagata, 2007) and *B. gymnorhiza* BURP Domain Containing protein3 (BgBDC3) (Banzai *et al.*, 2002b) (Fig. 2f). This suggests that salinity is an environmental factor that causes transcriptome variations in the natural environments. GO analysis of upregulated genes showed that genes involved in 'response to abscisic acid,' and 'response to water deprivation,' which are known to be related to salt-stress responses in plants (Hasegawa *et al.*, 2000; van Zelm *et al.*, 2020), were enriched in the upregulated genes in the saline environment (Fig. 2g; Tables S19–S22). On the contrary, many genes involved in chloroplast regulation were downregulated (Table S20), which may correlate with the decline in photosynthesis rate observed in the mangroves under salt and osmotic stresses (Miyama & Tada, 2008). In contrast, upregulated or downregulated genes in the control conditions only partially overlapped with those in the counterpart natural samples (Fig. S3b,c; Tables S23–S27). This may be due to natural or control condition-specific factors affecting gene expression profiles. Indeed, GO terms, 'response to chitin,' 'response to wounding,' 'response to salicylic acid,' and the KEGG pathway term 'Plant–pathogen interaction' were specifically enriched in the upregulated genes in the natural saline condition (Fig. 2g; Tables S19, S21), suggesting that pathogen challenge may also be one of the major environmental factors affecting transcriptome dynamics of *B. gymnorhiza* trees in their natural environment.

DNA hypermethylation in the *B. gymnorhiza* genome under high salinity environments

To investigate the whole-genome cytosine methylation of *B. gymnorhiza* in different natural environments, we performed WGBS, using both natural and control samples. Two biological individual replicates from each condition (two of each oceanside and riverside for natural experiment and two of each saline and brackish for the control experiment; in total, eight samples) were used (Fig. S4a; Tables S28, S29).

To investigate the DNA methylation patterns in different genomic regions, we analyzed the methylation level in the gene bodies (from transcription start site, TSS, to transcription termination site, TTS) as well as flanking regions of 1 kb upstream from TSS and 1 kb downstream from TTS (Figs 3c,d, S4b,c). The methylation level was highest in the CG context, followed by CHG and then CHH in each gene region (Fig. S4a), which is consistent with the methylome studies in other species (Al-Harrasi *et al.*, 2018; Jiang *et al.*, 2019; Zhang *et al.*, 2019). In the natural saline environment, DNA cytosine methylation level was elevated, especially in the CHG and CHH contexts, which mainly occurred in TE sequences (Fig. 3a,b). The non-CG hypermethylation in TEs was recapitulated in the samples grown under saline condition in control experiments (Fig. S4b,c), demonstrating that salinity might be one of the primary factors inducing the hypermethylation in TEs in the *B. gymnorhiza* genome. The observed non-CG hypermethylation was most
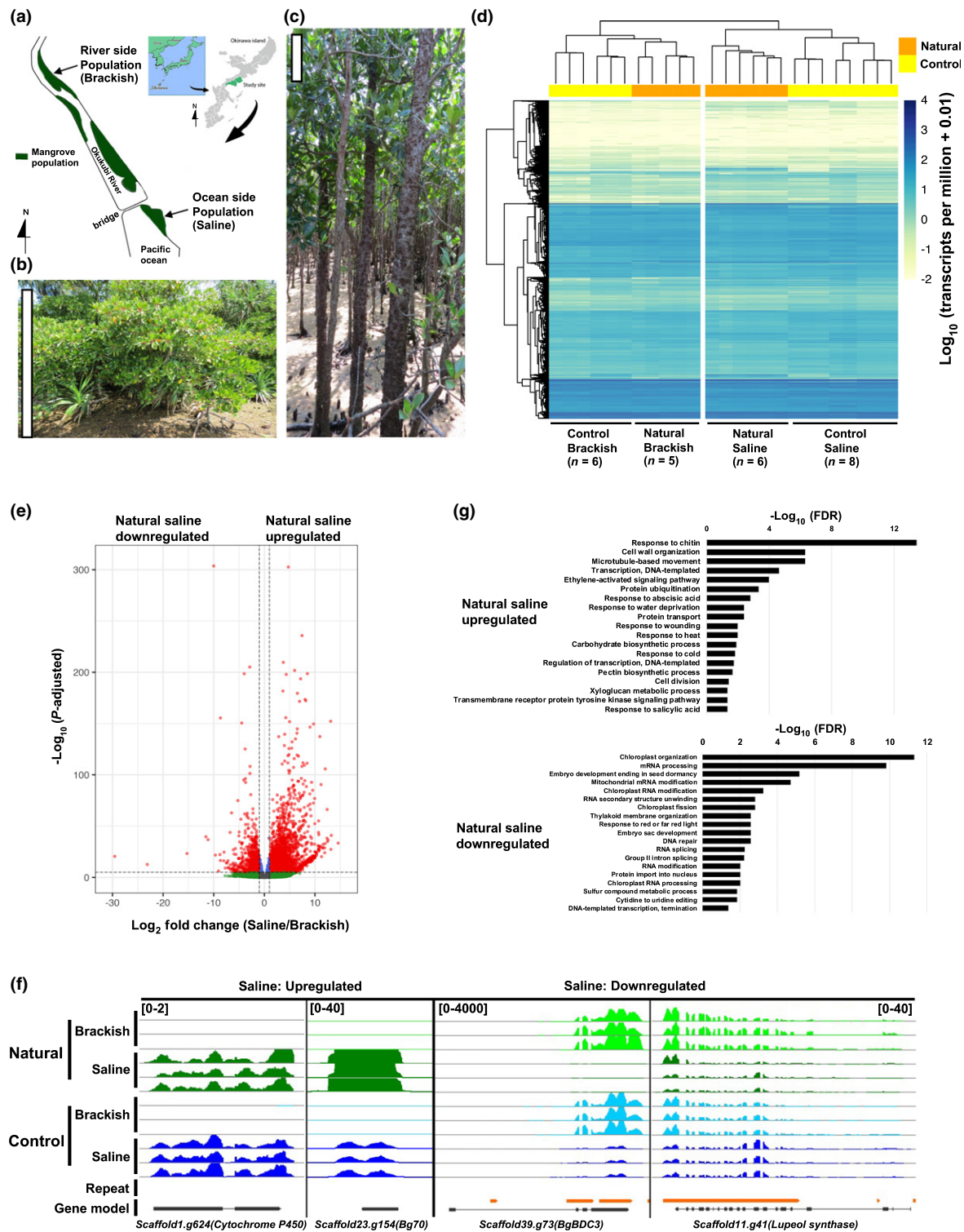
Fig. 2 (a) The location of the study site. Maps of Japan, the Okinawa main island, and the location of *Bruguiera gymnorhiza* populations: river side (brackish) and oceanside (saline), along the Okukubi River (26°27′N, 127°56′E). (b, c) Representative *B. gymnorhiza* trees in the river side population (b), and the oceanside population (c). Bars represent 100 cm. (d) A heat map showing the expression levels of *B. gymnorhiza* genes (*n* = 25 250) clustered by transcripts per million (TPM) of each gene (rows) and replicates (columns). The number of replicates for RNA-sequencing (RNA-Seq) in each condition is shown at the bottom of the heatmap. (e) Volcano plot showing differential gene expression in the natural environment. Values of log₂ fold change (saline/brackish) and adjusted *P*-value for genes were obtained by RNA-Seq data using DESEQ2 (Supporting Information Tables S17, S18). Vertical dashed lines indicate thresholds of ± 1 (0.5- and 2-fold changes), and the horizonal dashed line indicates a threshold of five (adjusted *P* = 1e⁻⁵). (f) Browser view of representative *B. gymnorhiza* gene loci showing differential gene expression in both natural and control conditions. Numbers in brackets indicate scales of tracks in count per million (CPM). Three representative RNA-Seq replicates are shown in each condition. (g) Results of gene ontology enrichment analysis of genes upregulated (top) and downregulated (bottom) under saline condition in the natural environment.
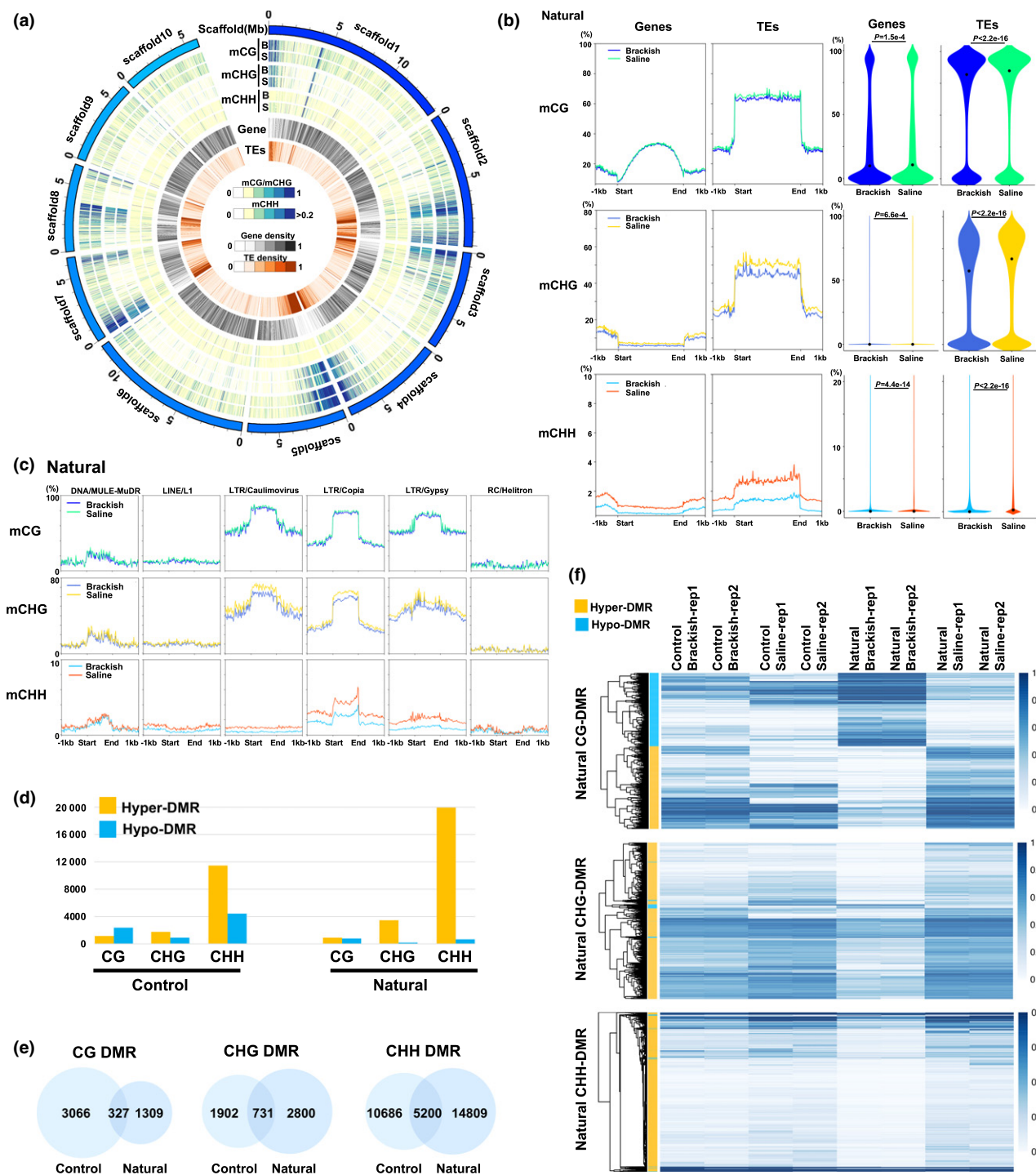
Fig. 3 (a) A Circos plot showing cytosine methylation levels of the 10 longest scaffolds (> 5 Mb) of *Bruguiera gymnorhiza*. The outermost circle shows the 10 scaffolds. The next six circles represent the DNA methylation levels in brackish (B) and saline (S) conditions for CG (0 to 1), CHG (0 to 1), and CHH (0 to 0.2) contexts (H can be A, or C, or T) in the natural environment. The first and second innermost circles show the density of genes (black) and transposable elements (TEs; orange), respectively. (b) (Left) Metaplots of cytosine methylation levels in CG, CHG, and CHH contexts for *genes* and TEs in brackish and saline conditions in the natural environment. (Right) Violin plots showing levels in CG, CHG, and CHH contexts for genes and TEs in brackish and saline conditions in the natural environment. Genes and TEs with ≥ 5 Cs were used for the analysis. Genes: mCG (brackish, $n = 31\,119$; saline, $n = 31\,293$); mCHG (brackish, $n = 32\,522$; saline, $n = 32\,648$); mCHH (brackish, $n = 33\,120$; saline, $n = 33\,155$). TEs: mCG (brackish, $n = 43\,020$; saline, $n = 44\,446$); mCHG (brackish, $n = 58\,549$; saline, $n = 60\,430$); mCHH (brackish, $n = 122\,428$; saline, $n = 124\,160$). Dots indicate median values. *P*-values by Brunner–Munzel test are indicated. (c) Metaplots of cytosine methylation levels in CG, CHG, and CHH contexts for TE families in brackish and saline conditions in the natural environment. (d) The number of differentially methylated regions (DMRs) in CG, CHG and CHH contexts in the control (left) and natural (right) experiments. (e) Venn diagrams indicating number of DMRs and overlap between DMRs of control and natural samples in CG, CHG and CHH contexts. (f) Heat maps of cytosine methylation levels in DMRs in natural samples and their corresponding regions in control samples, clustered by methylation levels (rows). Two replicates of BS-Seq data in indicated conditions are shown.

prominent in TE families with higher copy numbers, including LTR-Copia and LTR-Gypsy families (Figs 3c, S1d, S4c). To further study the differential methylation among different groups, we also identified the DMRs. The number of DMRs on each scaffold and the length distribution of DMRs are listed in Tables S30–S35. DMRs were identified by comparing the average methylation levels within a 100 bp window between saline and brackish, and those with statistical significance (FDR < 0.01) were used in the analysis. We detected 25 791 DMRs between oceanside and riverside natural samples and 21 918 DMRs between saline and brackish control samples. Many of the DMRs with hyper-CHH methylations were identified in the genomes of natural and control saline samples, which partially overlapped (Fig. 3d–f; Tables S30–S35), implying that other natural environmental factors besides salinity could affect the methylation variations among the samples from different environments.

The hyper-CHG and -CHH DMRs were mainly associated with TE sequences, while the majority of hyper- and hypo-CG DMRs were associated with genic regions in both natural and control samples (Figs 4a,b, S5a,b). We further examined the impact of changes in DNA methylation on the transcription of genes associated with DMRs. Genes associated with gene-body CG DMRs did not show significant differences between the hyper- and hypo-methylated gene groups (Figs 4c, S5c; Table S36). Similarly, genes associated with hyper-CHH DMRs in the promoter region (1 kb upstream of genes) or gene-body did not show significant changes in expression between saline and brackish samples in either natural or control environments, suggesting limited impacts of the observed DNA hypermethylation on the expression of the genes associated with DMRs in *B. gymnorhiza* leaf tissue.

Salinity-dependent regulation of BURP-domain containing protein gene cluster in the *B. gymnorhiza* genome

Although not many genes associated with DMRs showed expression changes, transcriptome analyses showed that there were several highly downregulated genes located on scaffold39 under the
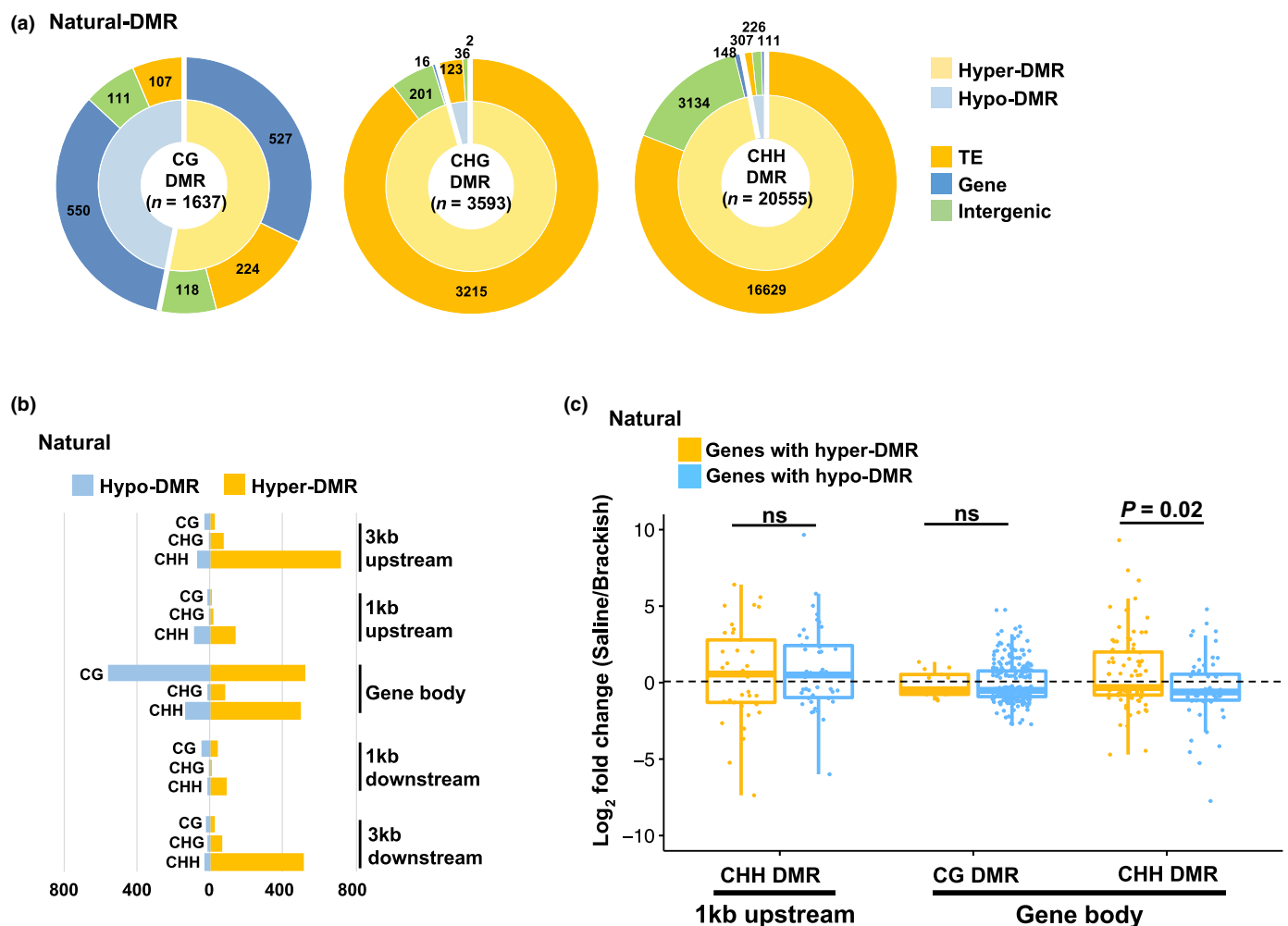

Fig. 4 (a) Pie charts showing genome features associated with natural-sample differentially methylated regions (DMRs). (b) Bar graphs representing the number of hypo- and hyper-DMRs in the genome features. (c) Box plots representing expression differences of genes (log$_2$ fold changes) associated with DMRs in the indicated genome features and contexts. Thick horizonal bars in the boxes indicate median values, box limits are upper and lower quartiles, and whiskers represent ×1.5 interquartile ranges. *P*-values by Brunner–Munzel test are indicated. ns, not significant (*P* > 0.05).

natural saline condition (Table S18). These genes encode BURP-domain containing proteins (Table S11), which were forming a cluster spanning the *c.* 300 kb region from 80 000 to 110 000 on scaffold39 (Fig. 5a). The BURP domain (named from protein families *B*NM2, *U*SP, *R*D22, *P*G1β) proteins (Hattori *et al.*, 1998) are involved in various plant metabolic pathways,

development, and stress responses (Wang *et al.*, 2015), and often form gene clusters in plant genomes (Li *et al.*, 2016). Previous studies on *B. gymnorhiza* identified BURP-domain containing protein genes *BgBDC1-4* (Banzai *et al.*, 2002b), which showed downregulation under high-salinity conditions (Fig. 2f) (Banzai *et al.*, 2002a; Miyama & Hanagata, 2007). We searched for
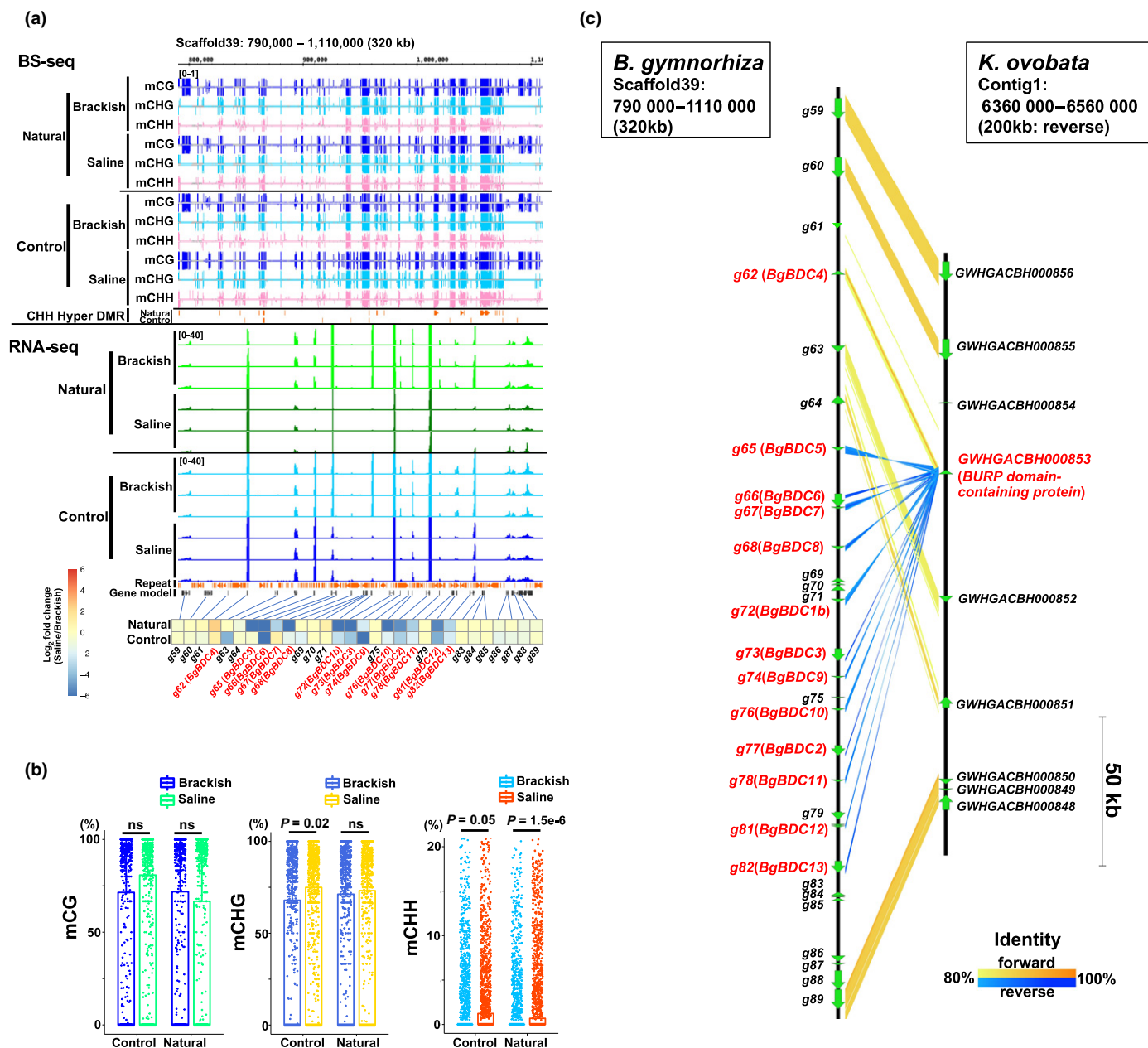

Fig. 5 (a) Browser view of the BURP gene cluster in the scaffold39 of the *Bruguiera gymnorhiza* genome. Tracks: top to bottom; cytosine methylation levels (0 to 1) in Watson and Crick strands in CG (blue), CHG (light blue), and CHH (pink) contexts (H can be A, or C, or T) of indicated conditions, CHH hyper-differentially methylated regions (DMRs; orange), RNA-sequencing tracks (count per million (CPM; 0 to 40)), repeats (orange), gene model (black), and heat maps showing differences in expression (log$_2$ fold changes) of genes, including BURP genes (*Bruguiera gymnorhiza* BURP-domain containing protein, BgBDCs; red). g80 and g81 in scaffold39 were originally annotated as two separate genes, while from the protein sequences they were likely from one BURP gene-coding sequences, and only expression of g81 is shown. (b) Boxplots showing cytosine methylation levels of the BURP gene cluster, from 800 201 to 1076 600 of scaffold39, averaged in 100-bp windows. *P*-values by Brunner–Munzel test are indicated. ns, not significant (*P* > 0.05). Thick horizonal bars in the boxes indicate median values, box limits are upper and lower quartiles, and whiskers represent ×1.5 interquartile ranges. (c) A synteny of the BURP gene cluster between *B. gymnorhiza* (scaffold39: 790 000–1100 000) and *Kandelia obovata* (Contig1: 6360 000–6560 000, reverse orientation). BURP genes are shown in red. BLAST identity (%) is indicated by ribbons (sense orientation, light blue to blue; reverse orientation, yellow to orange).

proteins homologous to BgBDCs in the *B. gymnorhiza* genome, identifying at least 14 BgBDC genes, including those highly similar to BgBDC1-4 genes (Fig. S6a). Among the 14 BgBDC genes, 13 genes were located in the cluster on the scaffold39 (Fig. S6a, b). In addition to BgBDC genes, many TEs and other repeats were accumulated within the cluster (Fig. 5a), which were highly methylated at both CG and non-CG contexts (Fig. 5a). These TEs were hypermethylated at CHH contexts under saline conditions in both natural and control environments (Fig. 5b), which could be associated with the cluster-wide gene regulation under salt stress. We further searched the syntenic regions of the cluster in the genome of *K. obovata* (Hu *et al.*, 2020), and found a syntenic block in contig1 of the *K. obovata* genome (Fig. 5c). Interestingly, only a single BURP domain protein gene is encoded in the corresponding region of *K. obovata,* suggesting either duplications and expansions of the BURP gene cluster in *B. gymnorhiza*, or contractions of the cluster in the *K. obovata* genome. Accumulation of the repetitive elements in the BURP gene cluster in the *B. gymnorhiza* genome might have contributed to the formation of the gene cluster, which has often been observed in other gene families in the plant genomes (Galindo-Gonzalez *et al.*, 2017; Lai & Eulgem, 2018).

Differential expression of chromatin modifier genes under salinity stresses in *B. gymnorhiza*

To understand the basis of the DNA hypermethylation of TEs in the *B. gymnorhiza* genome induced under saline conditions (Fig. 3), expression profiles of chromatin modifier genes in *B. gymnorhiza* were further investigated by searching the counterparts of *A. thaliana* (Arabidopsis Genome, 2000). Most of the homologs involved in DNA methylation pathways, including DNA methylases and demethylases, RNAi machinery, factors involved in RdDM, and histone H3K9 methylases and demethylases, were identified (Table S37), suggesting the conservations of the basic DNA methylation machineries in the *B. gymnorhiza* genome (Fig. 6a). Interestingly, the saline conditions in both natural and control environments significantly altered the expression of several genes involved in DNA methylation (Fig. 6a,b). In particular, the *B. gymnorhiza* homologs of the non-CG methylase genes *Bg_CMT2* and *Bg_CMT3* were significantly upregulated in the natural saline environment. These genes are required for CHG and CHH methylation of TEs, depending on H3K9 methylation in other plant species (Law & Jacobsen, 2010; Stroud *et al.*, 2013; Zemach *et al.*, 2013). In addition, the *de novo* methyltransferase homolog *Bg_DRM2*, which is involved in RdDM in plants (Matzke & Mosher, 2014; Erdmann & Picard, 2020) was significantly upregulated in both natural and control saline conditions (Figs 6a, S7). Conversely, one of the DNA demethylase homologs (Zhang *et al.*, 2018), *Bg_DME2,* showed rather significant downregulation under saline conditions (Fig. 6a,b). Although the degree of DNA methylation differences induced by the gene expression changes remains unclear, the upregulation of the non-CG methylase genes and downregulation of the DNA demethylase gene under saline conditions were consistent with the observed DNA hypermethylation of TEs,

especially in CHG and CHH sites in saline conditions. Taken together, this indicates that the differences in the expression of chromatin modifier genes may play a role in epigenome regulation in plants under high-salinity environments.

Discussion

TEs can contribute to genome expansion and evolution (Bennetzen & Wang, 2014), and are often activated by environmental stresses in plants (Ito *et al.*, 2011; Galindo-Gonzalez *et al.*, 2017; Lanciano & Mirouze, 2018). The genome of mangroves is the smallest among tree species (Xu *et al.*, 2017; Hu *et al.*, 2020). Previous comparative genome studies have reported that mangrove genomes show a convergent reduction of TE sequences compared to nonmangrove genomes (Lyu *et al.*, 2018), which can partly be attributed to RdDM pathway that suppresses the accumulation of young retrotransposons, such as Gypsy family, in the genome (Wang et al., 2018). The RdDM pathway primarily directs non-CG methylation at the LTRs of retrotransposons at the edges of the TE units (Stroud *et al.*, 2013; Zemach *et al.*, 2013). In contrast, CMTs mainly methylate non-CG sites in the body of TE sequences (Zemach *et al.*, 2013), and the convergence of RdDM and CMT pathways are required for stable silencing of TEs in plant genomes (Law & Jacobsen, 2010). The non-CG hypermethylation of TEs in the *B. gymnorhiza* genome under salt stress seems to occur at both the edges and the body of TE sequences (Figs 3, S4), concurrent with the activation of CMT and DRM methylases (Fig. 6), suggesting that both RdDM and CMT-mediated DNA methylation pathways may become active in *B. gymnorhiza* under saline conditions. The results suggest the presence of robust TE suppression mechanisms involving DNA methylation in the mangrove genome under salinity stress (Fig. 3). However, the *B. gymnorhiza* genome is still comparatively large, with a high TE proportion compared to other mangrove species. Among the mangrove family Rhizophoraceae, which diverged from nonmangrove species (at *c.* 56 Myr) (Guo *et al.*, 2017; Xu *et al.*, 2017), *B. gymnorhiza* diverged earlier, and its phylogenetic position was thus isolated from other mangroves, despite the monophyly of the *Rhizophoreae* tribe (Fig. 1) (Schwarzbach & Ricklefs, 2000; Guo *et al.*, 2017). *Bruguiera gymnorhiza* might regulate methylation of TEs less stringently as compared to other Rhizophoraceae species that diverged more recently; this would explain the differences in TE accumulation and the genome size differences.

TE sequences in the genome often enhance rearrangement, duplication, and recombination, which contribute to gene evolution and genome diversification (Bennetzen & Wang, 2014). The presence of many TEs in the BURP gene cluster (Fig. 5) suggested that the TEs might have contributed to cluster formation in *B. gymnorhiza*, as observed in the R-gene clusters essential for various pathogen responses in plant genomes (Galindo-Gonzalez *et al.*, 2017; Lai & Eulgem, 2018). Indeed, we found intron-less BURP genes (*BgBDC7*, *BgBDC9*, *BgBDC10*, *BgBDC11*) scattered in the cluster, suggesting previous retroposition or gene transduction events that are often triggered by the presence of nearby LTR retrotransposon sequences (Galindo-Gonzalez *et al.*, 2017). Thus, there may be a trade-off between the cost of
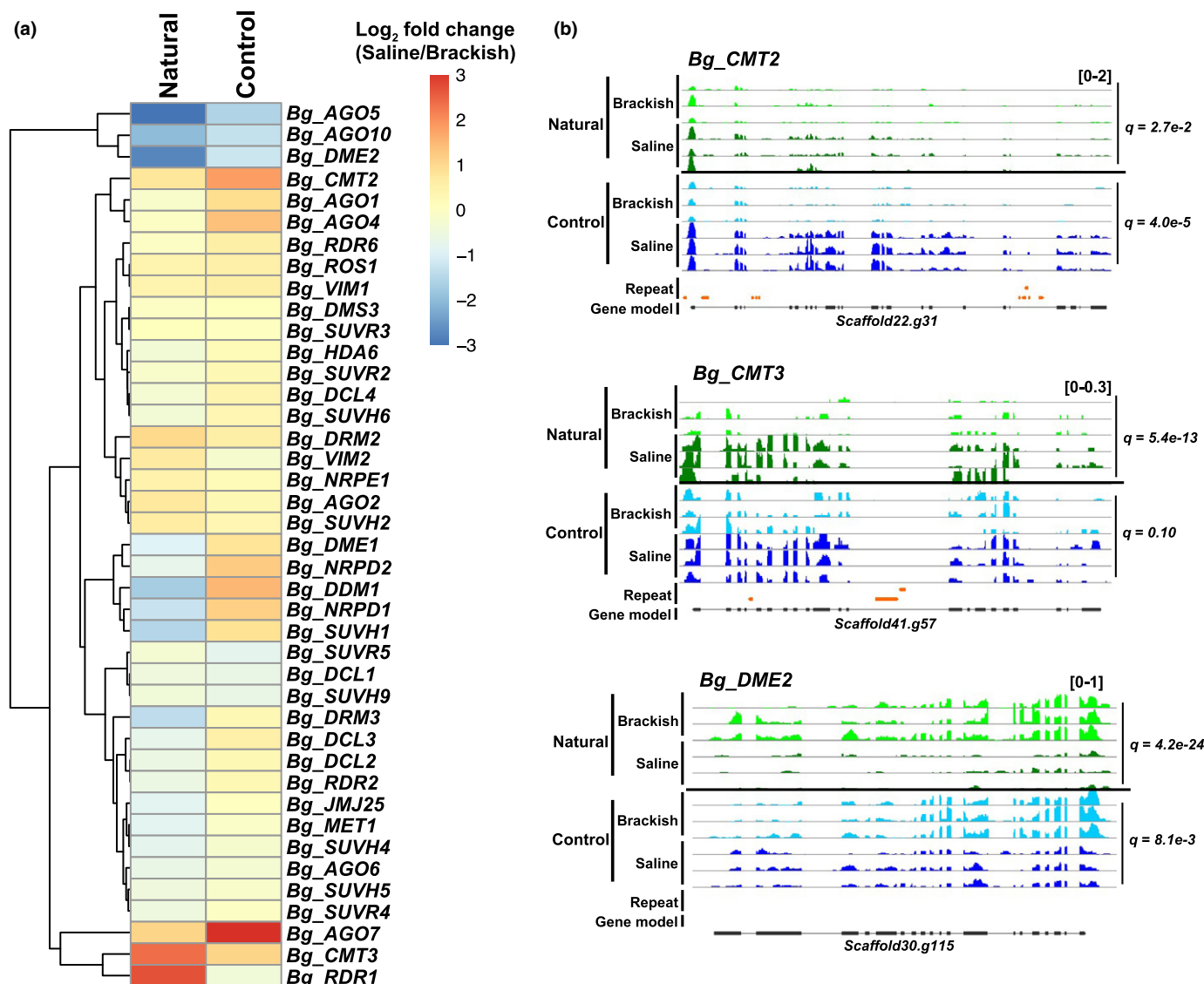
Fig. 6 (a) Heat map showing differences in expression (log$_2$ fold changes (saline/brackish)) of genes homologous to chromatin modifiers. The dendrogram shows hierarchical gene clustering by expression levels in natural and control conditions (rows). (b) Genome browser view of three representatives of chromatin modifier genes showing differential expression in both natural and control conditions. Numbers in brackets indicate scales of tracks in count per million (CPM). Three representative RNA-sequencing replicates are shown in each condition. q-Values (adjusted P-values) calculated by DE$_{SEQ2}$ are indicated.

maintaining large genomes under harsh environments and the accumulations of TEs in the genomes as a source of genetic diversity. Epigenetic control of TEs could also contribute to the differentiation of nearby gene expression and phenotypic plasticity, which may play a role in the adaptive response of mangrove trees to their stressful environment. Indeed, mangrove trees, including *B. gymnorhiza*, often show drastic morphological changes in responses to the environmental stresses (Fig. 2) (Suwa *et al.*, 2009; Lira-Medeiros *et al.*, 2010), suggesting phenotypic and physiological adaptations to the harsh growing environments. Intriguingly, our transcriptome analysis revealed a significant downregulation of the RNAi component *ARGONAUTE10* (*AGO10*) (Fig. S7), which is involved in stem cell maintenance via miRNA regulation in *A. thaliana*, and the mutation of

AGO10 causes stem cell termination in the shoot apical meristem (Liu *et al.*, 2009; Zhu *et al.*, 2011). The downregulation of the *AGO10* could be a factor leading to the dwarf phenotypes of *B. gymnorhiza* under saline conditions (Fig. 2). In addition, a recent finding of cell cycle-coupled regulation of DNA methylation (Borges *et al.*, 2021) suggests that differences in cell cycle activities in *B. gymnorhiza* plants growing under different environmental conditions might affect DNA methylation levels in the genome. The observed morphological diversification of the mangrove tree (Fig. 2) might be a result of adaptive control of developmental growth in the saline conditions rather than passive responses to environmental stresses, which may allow the allocation of limited resources to other essential cellular processes.

Differential DNA methylation pattern has been known to have an impact in regulation of salt responsive genes. For example, studying salt-treated palm seedlings has shown that salinity stress causes insufficient root growth and lower expression of some key photosynthetic genes, which is suggested to be due to DNA methylation changes (Sperling *et al.*, 2014; Al-Harrasi *et al.*, 2018). Other studies on *Arabidopsis* (Boyko *et al.*, 2010), alfalfa (*Medicago* spp.) (Al-Lawati *et al.*, 2016), and *Mesembryanthemum crystallinum* (Dyachenko *et al.*, 2006) have all reported DNA methylation to be associated with salinity tolerance phenotype. Salt-sensitive rice cultivars that experienced salt stress showed changes in DNA methylation where hypomethylation caused by salinity was correlated with upregulation of DNA demethylase genes (Ferreira *et al.*, 2015). Non-CG methylation of a salt tolerance gene in *Arabidopsis* also showed to help with the adaptation to high-salinity stress (Ashapkin *et al.*, 2020). A study on wheat has shown a similar result, where hyper-methylation caused by salt stress was correlated with the regulation of salt-responsive genes (Kumar *et al.*, 2017). Similar to our results, DNA methylation through the activity of methyltransferases such as MET1, CMT3, and DRM2 has been reported extensively in response to stress (Shen *et al.*, 2014; Naydenov *et al.*, 2015). Thus, the observed global changes in DNA methylation in *B. gymnorhiza* under saline conditions may be a conserved response in various plant species.

In the study site of mangrove forest, *B. gymnorhiza* showed the greatest morphological differences between the oceanside and riverside compared to the other two neighboring mangrove species *R. stylosa* and *K. obovata*. Indeed, *Bruguiera* has several unique characteristics that distinguish it from other mangrove species. For example, seedling dispersal in *Bruguiera* is accompanied with the fruits, whereas only seedlings are dispersed in other mangroves (Juncosa, 1984; Tomlinson, 1986). Furthermore, *B. gymnorhiza* is mainly pollinated by birds, while *Rhizophora* and other close relatives are wind- and insect-pollinated (Kondo *et al.*, 1987). Additionally, *B. gymnorhiza* has an extraordinarily wide distribution range throughout the Indo-West Pacific from East Africa to Southeast Asia, Australia, and the western Pacific (Duke & Ge, 2011). Future studies of various mangrove species would allow us to better understand the genome strategies of plant species for adaptation to extreme natural environments.

Conclusions

In this study, we performed a *de novo* genome assembly and *in natura* epigenomics of the mangrove tree *B. gymnorhiza*. The results revealed that *B. gymnorhiza* under high-salinity conditions induced DNA hypermethylation in TEs, suggesting a robust epigenome regulation for TEs in *B. gymnorhiza* under salt stresses. The high-quality reference genome, transcriptome, and epigenome data in this study builds a promising foundation for future genetic and epigenetic studies in mangroves and may also contribute to a better understanding of the adaptive responses and epigenome regulation in plant species under fluctuating environmental conditions. Our results support the involvement of DNA methylation in some differentiation of gene expression and phenotypic characteristics that may be part of the adaptive response of mangrove trees to their stressful environment. Future studies are needed to understand the mechanism behind this connection.

Author contributions

The study was conceived and coordinated by MM and HS, all experiments were designed and performed by MM. For data and bioinformatic analyses, flow cytometry and *k*-mer analyses were performed by MM. Short read genome assembly was performed by MM, PacBio genome assembly was performed by MM and FM, scaffolding and gap closing were performed by MM, annotation of the genome was performed by MM, FM, and was assisted by DG, Transcriptome assembly was performed by MM and FM, Gene expression analysis was performed by MM and was assisted by HS and DG, DNA methylation analyses were performed by MM and HS. Figures and tables were prepared by MM and HS. The manuscript was written by MM and HS.

ORCID

Daria Gavriouchkina (iD) https://orcid.org/0000-0002-7488-5898
Ferdinand Marlétaz (iD) https://orcid.org/0000-0001-8124-4266
Matin Miryeganeh (iD) https://orcid.org/0000-0002-5685-9223
Hidetoshi Saze (iD) https://orcid.org/0000-0002-8647-1430

Data availability

The *B. gymnorhiza* genome data, as well as DNA-Seq, RNA-Seq, and BS-Seq data, are available in the NCBI BioProject under accession no. PRJNA703772. In-house code is available as Notes S2.

References

Al-Harrasi I, Al-Yahyai R, Yaish MW. 2018. Differential DNA methylation and transcription profiles in date palm roots exposed to salinity. *PLoS ONE* **13**: e0191492.

Al-Lawati A, Al-Bahry S, Victor R, Al-Lawati AH, Yaish MW. 2016. Salt stress alters DNA methylation levels in alfalfa (*Medicago* spp). *Genetics and Molecular Research* **15**: 15018299.

Arabidopsis Genome I. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Ashapkin VV, Kutueva LI, Aleksandrushkina NI, Vanyushin BF. 2020. Epigenetic mechanisms of plant adaptation to biotic and abiotic stresses. *International Journal of Molecular Sciences* 21: 7457.

Banzai T, Hershkovits G, Katcoff DJ, Hanagata N, Dubinsky Z, Karube I. 2002a. Identification and characterization of mRNA transcripts differentially expressed in response to high salinity by means of differential display in the mangrove, *Bruguiera gymnorrhiza*. *Plant Science* 162: 499–505.

Banzai T, Sumiya K, Hanagata N, Dubinsky Z, Karube I. 2002b. Molecular cloning and characterization of genes encoding BURP domain-containing protein in the mangrove, *Bruguiera gymnorrhiza*. *Trees–Structure and Function* 16: 87–93.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58: 268–276.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 57: 289–300.

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology* 65: 505–530.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27: 578–579.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Borges F, Donoghue MTA, LeBlanc C, Wear EE, Tanurdžić M, Berube B, Brooks A, Thompson WF, Hanley-Bowdoin L, Martienssen RA. 2021. Loss of small-RNA-directed DNA methylation in the plant cell cycle promotes germline reprogramming and somaclonal variation. *Current Biology* 31: 591–600.e4.

Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollander J, Meins F Jr, Kovalchuk I. 2010. Transgenerational adaptation of arabidopsis to stress requires DNA methylation and the function of dicer-like proteins. *PLoS ONE* 5: e9514.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* 1962: 1–14.

Chen SF, Zhou RC, Huang YL, Zhang M, Yang GL, Zhong CR, Shi SH. 2011. Transcriptome sequencing of a highly salt tolerant mangrove species *Sonneratia alba* using Illumina platform. *Marine Genomics* 4: 129–136.

Cheong WH, Tan YC, Yap SJ, Ng KP. 2015. ClicO FS: an interactive web-based service of Circos. *Bioinformatics* 31: 3685–3687.

Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261: 201–217.

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM. 2009. Shedding light on an extremophile lifestyle through transcriptomics. *New Phytologist* 183: 764–775.

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM. 2010. Comparative transcriptomics for mangrove species: an expanding resource. *Functional & Integrative Genomics* 10: 523–532.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences, USA* 109: E2183–E2191.

Duke NC, Ge XJ. 2011. Bruguiera (Rhizophoraceae) in the Indo-West Pacific: a morphometric assessment of hybridization within single-flowered taxa. *Blumea* 56: 36–48.

Dyachenko OV, Zakharchenko NS, Shevchuk TV, Bohnert HJ, Cushman JC, Buryanov YI. 2006. Effect of hypermethylation of CCWGG sequences in DNA of *Mesembryanthemum crystallinum* plants on their adaptation to salt stress. *Biochemistry (Moscow)* 71: 461–465.

Erdmann RM, Picard CL. 2020. RNA-directed DNA methylation. *PLoS Genetics* 16: e1009034.

Feng X, Xu SH, Li JF, Yang YC, Chen QP, Lyu HM, Zhong CR, He ZW, Shi SH. 2020. Molecular adaptation to salinity fluctuation in tropical intertidal environments of a mangrove tree *Sonneratia alba*. *BMC Plant Biology* 20: 178.

Ferreira LJ, Azevedo V, Maroco J, Oliveira MM, Santos AP. 2015. Salt tolerant and sensitive rice varieties display differential methylome flexibility under salt stress. *PLoS ONE* 10: e0124060.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences, USA* 117: 9451–9457.

Freese NH, Norris DC, Loraine AE. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32: 2089–2095.

Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA. 2017. LTR-retrotransposons in plants: engines of evolution. *Gene* 626: 14–25.

Garrison E, Marth G. 2012. *Haplotype-based variant detection from short-read sequencing*. arXiv:1207.3907v2.

Ghelfi A, Shirasawa K, Hirakawa H, Isobe S. 2019. Hayai-annotation plants: an ultra-fast and comprehensive functional gene annotation system in plants. *Bioinformatics* 35: 4427–4429.

Góngora-Castillo E, Childs KL, Fedewa G, Hamilton JP, Liscombe DK, Magallanes-Lundback M, Mandadi KK, Nims E, Runguphan W, Vaillancourt B et al. 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE* 7: e52506.

Guo WX, Wu HD, Zhang Z, Yang C, Hu L, Shi XG, Jian SG, Shi SH, Huang YL. 2017. Comparative analysis of transcriptomes in Rhizophoraceae provides insights into the origin and adaptive evolution of mangrove plants in intertidal environments. *Frontiers in Plant Science* 8: 795.

Guo WL, Zhu P, Pellegrini M, Zhang MQ, Wang XF, Ni ZF. 2018. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* 34: 381–387.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072–1075.

Hao ZD, Lv DK, Ge Y, Shi JS, Weijers D, Yu GC, Chen JH. 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science* 6: e251.

Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ. 2000. Plant cellular and molecular responses to high salinity. *Annual Review of Plant Physiology and Plant Molecular Biology* 51: 463–499.

Hattori J, Boutilier KA, Campagne MMV, Miki BL. 1998. A conserved BURP domain defines a novel group of plant proteins with unusual primary structures. *Molecular and General Genetics* 259: 424–428.

Hu M-J, Sun W-H, Tsai W-C, Xiang S, Lai X-K, Chen D-Q, Liu X-D, Wang Y-F, Le Y-X, Chen S-M et al. 2020. Chromosome-scale assembly of the *Kandelia obovata* genome. *Horticulture Research* 7: 75.

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472: 115–119.

Jiang S-H, Sun Q-G, Chen M, Wang N, Xu H-F, Fang H-C, Wang Y-C, Zhang Z-Y, Chen X-S. 2019. Methylome and transcriptome analyses of apple fruit somatic mutations reveal the difference of red phenotype. *BMC Genomics* 20: 117.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21: 241.

Juncosa AM. 1984. Biology and ecology of mangroves. *Ecology* 65: 1698.

Kondo K, Nakamura T, Tsuruda K, Saito N, Yaguchi Y. 1987. Pollination in *Bruguiera gymnorrhiza* and *Rhizophora mucronata* (Rhizophoraceae) in Ishigaki Island, the Ryukyu Islands, Japan. *Biotropica* 19: 377–380.

Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. 2018. De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology* 36: 1174–1182.

Krueger F, Andrews SR. 2011. BISMARK: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572.

Kumar S, Beena AS, Awana M, Singh A. 2017. Salt-induced tissue-specific cytosine methylation downregulates expression of HKT genes in contrasting wheat (*Triticum aestivum* L.) genotypes. *DNA and Cell Biology* 36: 283–294.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.

Lai Y, Eulgem T. 2018. Transcript-level expression control of plant NLR genes. *Molecular Plant Pathology* 19: 1267–1281.

Lamke J, Baurle I. 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology* 18: 124.

Lanciano S, Mirouze M. 2018. Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Current Opinion in Genetics & Development* 49: 106–114.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* 11: 204–220.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li Y, Chen X, Chen Z, Cai RH, Zhang HM, Xiang Y. 2016. Identification and expression analysis of BURP domain-containing genes in *Medicago truncatula*. *Frontiers in Plant Science* 7: 485.

Liao Y, Smyth GK, Shi W. 2014. FEATURECOUNTS: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930.

Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PC. 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE* 5: e10326.

Liu QL, Yao XZ, Pi LM, Wang H, Cui XF, Huang H. 2009. The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in Arabidopsis. *The Plant Journal* 58: 27–40.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.

Lyu HM, He ZW, Wu CI, Shi SH. 2018. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *New Phytologist* 217: 428–438.

Mapleson D, Venturini L, Kaithakottil G, Swarbreck D. 2018. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *Gigascience* 7: 131.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770.

Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* 15: 394–408.

Meera SP, Augustine A. 2020. De novo transcriptome analysis of *Rhizophora mucronata* Lam. furnishes evidence for the existence of glyoxalase system correlated to glutathione metabolic enzymes and glutathione regulated transporter in salt tolerant mangroves. *Plant Physiology and Biochemistry* 155: 683–696.

Miura F, Enomoto Y, Dairiki R, Ito T. 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research* 40: e136.

Miyama M, Hanagata N. 2007. Microarray analysis of 7029 gene expression patterns in Burma mangrove under high-salinity stress. *Plant Science* 172: 948–957.

Miyama M, Shimizu H, Sugiyama M, Hanagata N. 2006. Sequencing and analysis of 14,842 expressed sequence tags of Burma mangrove, *Bruguiera gymnorrhiza*. *Plant Science* 171: 234–241.

Miyama M, Tada Y. 2008. Transcriptional and physiological study of the response of Burma mangrove (*Bruguiera gymnorhiza*) to salt and osmotic stress. *Plant Molecular Biology* 68: 119–129.

Nawrocki EP, Eddy SR. 2013. INFERNAL 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29: 2933–2935.

Naydenov M, Baev V, Apostolova E, Gospodinova N, Sablok G, Gozmanova M, Yahubyan G. 2015. High-temperature effect on genes engaged in DNA methylation and affected by DNA methylation in Arabidopsis. *Plant Physiology and Biochemistry* 87: 102–108.

Niknafs YS, Pandian B, Iyer HK, Chinnaiyan AM, Iyer MK. 2017. TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nature Methods* 14: 68–70.

Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jönsson KA, Zhou QI et al. 2021. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources* 21: 263–286.

Raju Aluri JS. 2014. Reproductive ecology of mangrove flora: conservation and management. *Transylvanian Review of Systematical and Ecological Research* 15: 133–184.

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. DEEPTOOLS2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* 44: W160–W165.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GENOMESCOPE 2.0 and SMUDGEPLOT for reference-free profiling of polyploid genomes. *Nature Communications* 11: 1432.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16: 276–277.

Sahu PP, Pandey G, Sharma N, Puranik S, Muthamilarasan M, Prasad M. 2013. Epigenetic mechanisms of plant stress responses and adaptation. *Plant Cell Reports* 32: 1151–1159.

Schwarzbach AE, Ricklefs RE. 2000. Systematic affinities of Rhizophoraceae and Anisophylleaceae, and intergeneric relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. *American Journal of Botany* 87: 547–564.

Shen X, De Jonge J, Forsberg SK, Pettersson ME, Sheng Z, Hennig L, Carlborg Ö. 2014. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genetics* 10: e1004842.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* 8: 272–285.

Soderlund C, Nelson W, Shoemaker A, Paterson A. 2006. SYMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Research* 16: 1159–1168.

Soubry A. 2015. Epigenetic inheritance and evolution: a paternal perspective on dietary influences. *Progress in Biophysics and Molecular Biology* 118: 79–85.

Sperling O, Lazarovitch N, Schwartz A, Shapira O. 2014. Effects of high salinity irrigation on growth, gas-exchange, and photoprotection in date palms (*Phoenix dactylifera* L., cv. Medjool). *Environmental and Experimental Botany* 99: 100–109.

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152: 352–364.

Sullivan MJ, Petty NK, Beatson SA. 2011. EASYFIG: a genome comparison visualizer. *Bioinformatics* 27: 1009–1010.

Suwa R, Deshar R, Hagihara A. 2009. Forest structure of a subtropical mangrove along a river inferred from potential tree height and biomass. *Aquatic Botany* 91: 99–104.

Thankachan SV, Chockalingam SP, Liu YC, Apostolico A, Aluru S. 2016. ALFRED: a practical method for alignment-free distance computation. *Journal of Computational Biology* 23: 452–460.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.

Tomlinson P. 1986. *The botany of mangroves*. Cambridge, UK: Cambridge University Press.

van Zelm E, Zhang YX, Testerink C. 2020. Salt tolerance mechanisms of plants. *Annual Review of Plant Biology* 71: 403–433.

Wang LH, Wu NN, Zhu Y, Song WL, Zhao X, Li YX, Hu YK. 2015. The divergence and positive selection of the plant-specific BURP-containing protein family. *Ecology and Evolution* 5: 5394–5412.

Wang Y, Liang W, Tang T. 2018. Constant conflict between Gypsy LTR retrotransposons and CHH methylation within a stress-adapted mangrove genome. *New Phytologist* 220: 922–935.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S et al. 2007. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 35: D5–D12.

Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J et al. 2016. Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* 5: e13546.

Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y et al. 2017. The origin, diversification and adaptation of a major mangrove

clade (Rhizophoreae) revealed by whole-genome sequencing. *National Science Review* 4: 721–734.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* 35: W265–W268.

Yang YC, Yang SH, Li JF, Deng YF, Zhang Z, Xu SH, Guo WX, Zhong CR, Zhou RC, Shi SH. 2015. Transcriptome analysis of the Holly mangrove *Acanthus ilicifolius* and its terrestrial relative, *Acanthus leucostachyus*, provides insights into adaptation to intertidal zones. *BMC Genomics* 16: 605.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The Arabidopsis Nucleosome Remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153: 193–205.

Zhang H, Lang Z, Zhu JK. 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology* 19: 489–506.

Zhang M, Zhang X, Guo L, Qi T, Liu G, Feng J, Shahzad K, Zhang B, Li X, Wang H *et al.* 2019. Single-base resolution methylome of cotton cytoplasmic male sterility system reveals epigenomic changes in response to high-temperature stress during anther development. *Journal of Experimental Botany* 71: 951–969.

Zhao S, Li CI, Guo Y, Sheng Q, Shyr Y. 2018. RnaSeqSampleSize: real data based sample size estimation for RNA sequencing. *BMC Bioinformatics* 19: 191.

Zhu HL, Hu FQ, Wang RH, Zhou X, Sze SH, Liou LW, Barefoot A, Dickman M, Zhang XR. 2011. Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* 145: 242–256.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 The *Bruguiera gymnorhiza* genome and annotations.

Fig. S2 The *Bruguiera gymnorhiza* chloroplast genome with gene annotation.

Fig. S3 The expression analysis of *Bruguiera gymnorhiza* genes in Natural and Control conditions.

Fig. S4 DNA methylation analysis of *Bruguiera gymnorhiza* genome in Natural and Control conditions.

Fig. S5 The expression of *Bruguiera gymnorhiza* genes associated with DMRs in the Control condition.

Fig. S6 Protein analysis of BURP genes in the *Bruguiera gymnorhiza* genome.

Notes S1 Supplementary details about Materials and Methods.

Notes S2 In-house PYTHON script to obtain exon hints from transcriptome.

Table S1 Raw and Trimmed Illumina DNA sequencing information used in this study.

Table S2 PACBIO DNA sequencing information used in this study.

Table S3 Genome and transcriptome assembly results.

Table S4 Quality control of final assembly by mapping Hi-Seq data to final assembly.

Table S5 BUSCO analysis of the genome and the potential proteins.

Table S6 Repeat annotations.

Table S7 The intersect between LTR-finder and REPEATMODELER output.

Table S8 Top BLAST hits for genes based on a repeat-masked genome against Uniprot plant database.

Table S9 Synteny blocks between *Bruguiera gymnorhiza* and *Kandelia obovata*, *B. gymnorhiza* and *Rhizophora apiculata*, and *K. obovata* and *R. apiculate*.

Table S10 RNA-Seq mapping rate by STAR_ASTERIX.

Table S11 34 403 Protein-coding gene models in the *Bruguiera gymnorhiza* genome.

Table S12 tRNA annotation.

Table S13 snoRNA_miRNA_rRNA annotations.

Table S14 Repeat_summary.

Table S15 BLAST hits to *Bruguiera gymnorriza* ESTs.

Table S16 Raw Illumina RNA sequencing information used in this study.

Table S17 Upregulated genes in oceanside (saline) samples.

Table S18 Downregulated genes in oceanside (saline) samples.

Table S19 GO enrichment in biological process for upregulated genes in oceanside samples.

Table S20 GO enrichment in biological process for downregulated genes in oceanside samples.

Table S21 KEGG pathway enrichment for upregulated genes in oceanside samples.

Table S22 KEGG pathway enrichment for downregulated genes in oceanside samples.

Table S23 Upregulated genes in saline control condition.

Table S24 Downregulated genes in saline control samples.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.

- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <23 days. There are **no page or colour charges** and a PDF version will be provided for each article.

- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.

- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)

- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**