Research

New Phytologist

Genomic insights into molecular adaptation to intertidal environments in the mangrove *Aegiceras corniculatum*

Xiao Feng[1]*, Guohong Li[1]*, Shaohua Xu[1], Weihong Wu[1], Qipian Chen[1], Shao Shao[1], Min Liu[1], Nan Wang[1], Cairong Zhong[2], Ziwen He[1] and Suhua Shi[1]

[1]State Key Laboratory of Biocontrol, Guangdong Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; [2]Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou 571100, China

Author for correspondence:
Ziwen He
Email: heziwen@mail.sysu.edu.cn

Received: *11 March 2021*
Accepted: *2 June 2021*

New Phytologist (2021) **231:** 2346–2358
doi: 10.1111/nph.17551

Key words: *Aegiceras corniculatum*, crypto-vivipary, homeostasis, mangrove, molecular adaptation, whole-genome sequencing.

Summary

- Mangroves have colonised extreme intertidal environments characterised by high salinity, hypoxia and other abiotic stresses. *Aegiceras corniculatum*, a pioneer mangrove species that has evolved two specialised adaptive traits (salt secretion and crypto-vivipary) is an attractive ecological model to investigate molecular mechanisms underlying adaptation to intertidal environments.
- We assembled *de novo* a high-quality reference genome of *A. corniculatum* and performed comparative genomic and transcriptomic analyses to investigate molecular mechanisms underlying adaptation to intertidal environments.
- We provide evidence that *A. corniculatum* experienced a whole-genome duplication (WGD) event *c.* 35 Ma. We infer that maintenance of cellular environmental homeostasis is an important adaptive process in *A. corniculatum*. The 14-3-3 and H^+-ATPase protein-coding genes, essential for the salt homeostasis, were preferentially retained after the recent WGD event. Using comparative transcriptomics, we show that genes upregulated under high-salt conditions are involved in salt transport and ROS scavenging. We also found that all homologues of *DELAY OF GERMINATION1* (*DOG1*) had lost their heme-binding ability in *A. corniculatum*, and that this may contribute to crypto-vivipary.
- Our study provides insight into the genomic correlates of phenotypic adaptation to intertidal environments. This could contribute not only within the genomics community, but also to the field of plant evolution.

Introduction

Organismal adaptation is a fundamental topic in evolutionary biology. Intertidal zones are extreme environments that constantly experience drastic changes, with high and fluctuating salinity, hypoxia and other abiotic stresses (Giri *et al.*, 2011). Mangroves have colonised and have become well adapted to these habitats, evolving a series of highly specialised traits including salt tolerance, viviparous embryos, high tannin content and aerial roots (Liang *et al.*, 2008; Parida & Jha, 2010; Tomlinson, 2016). Therefore, they are attractive ecological model systems to investigate molecular mechanisms underlying adaptation to intertidal environments.

Vivipary and high salt tolerance are particularly common in mangroves and rare in other woody plants. Salt tolerance is a long-term and dynamic process influenced by multiple genes involving many morphological, physiological, molecular and cellular processes (Munns & Tester, 2008; Feng X *et al.*, 2020b). Several important signalling pathways have been identified, especially the salt overly sensitive (SOS) signalling pathway, as well as Ca^{2+}-dependent and ABA signalling pathways (Ji *et al.*, 2013). These systems can maintain and regulate cellular environmental homeostasis through mediating cellular signalling under salt stress. Vivipary in flowering plants is defined as the offspring's continuous growth when still attached to the mother plant without an apparent dormant period. Embryos of viviparous species, for example *Rhizophora*, *Bruguiera*, *Ceriops* and *Kandelia* mangroves, first break through the seed coat and then out of the fruit wall. By contrast, crypto-viviparous mangrove species *Aegiceras*, *Avicennia, Aegialitis* and *Nypa* only break out of the seed coat but not the fruit wall before dehiscence (Elmqvist & Cox, 1996; Shi *et al.*, 2005; Tomlinson, 2016). Previous studies have suggested that viviparous propagules are protected from high salinity and other stresses during early development (Zheng *et al.*, 1999; Liang *et al.*, 2008). These strategies can enhance mangrove reproductive potential in an unstable and extreme environment.

Aegiceras corniculatum, one of the most pervasive mangrove species widespread in the Indo West Pacific (IWP) region, is regarded as a pioneer mangrove species for its strong salt tolerance, usually growing along estuary banks (Supporting information Fig. S1). It is a typical outer seaward mangrove (Tomlinson,

*These authors contributed equally to this work.

2016; He *et al.*, 2019). *Aegiceras corniculatum* has developed two specialised adaptive traits: salt secretion glands and cryptovivipary (Ball, 1988; Ge & Sun, 1999). The lack of 'omics data for *A. corniculatum* is a major obstacle to understanding the genomics of acquisition and maintenance of these specialised traits and therefore this species' adaptation to intertidal environments. In this study, we report a high-quality reference genome of *A. corniculatum*. Through comparative and evolutionary analyses, we attempted to understand the genomic correlates of phenotypic adaptation to intertidal environments. The genome sequence provided here will accelerate genomic and evolutionary studies of this mangrove species and others.

Materials and Methods

Plant material

We sampled with permission one mature *Aegiceras corniculatum* plant from the nursery of Dongzhai Harbor National Nature Reserve in Hainan. Fresh and healthy leaves were harvested and immediately frozen in liquid nitrogen, followed by preservation at −80°C in the laboratory before DNA extraction. High-quality genomic DNA was extracted from leaves using the modified cetyltrimethylammonium bromide (CTAB) method. RNase A was used to remove RNA contaminants. The quality of the extracted DNA was examined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The quantity was estimated by electrophoresis on a 0.8% agarose gel. Total RNA from leaves was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's protocol.

PacBio long-read library preparation and sequencing

Single-Molecule Real-Time (SMRT) long-read sequencing was performed on a PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA). A single SMRT-bell library with 40-kb long inserts was constructed from sheared genomic DNA using a template library preparation workflow. The library was sequenced on PacBio SMRT cells 8 M (acquiring one movie of 15 h per SMRT cell) using a PacBio Sequel II instrument. After data filtering and preprocessing, 10.24 million long reads were generated, yielding *c.* 139.72 Gb (169× coverage) with an average read length of 13 647 bp.

Short-read sequencing

For DNA short-read sequencing, 150-bp paired-end libraries were prepared for sequencing on an Illumina NovaSeq 6000 platform and yielded *c.* 54.71 Gb of bases. The RNA-seq library was sequenced on a BGI-seq 500 platform for gene prediction and yielded *c.* 8.06 Gb of bases.

De novo genome assembly

We assembled the *A. corniculatum de novo* genome based on the PacBio long reads using three assemblers: Canu (Koren *et al.*, 2017), Falcon (Chin *et al.*, 2016) and Mecat2 (Xiao *et al.*, 2017) with optimised parameters. The optimal assembly (using Mecat2) was further polished with Racon (v.1.3.1) using long reads (Vaser *et al.*, 2017). To improve primary assembly accuracy, we corrected the remaining errors using Pilon (v.1.22) based on Illumina short reads (Walker *et al.*, 2014). The Purge Haplotigs pipeline was used to filter redundant sequences due to heterozygosity (Roach *et al.*, 2018).

Genome annotations

We identified repetitive sequences in the *A. corniculatum* genome by integrating homology-based and *de novo* approaches. For homology-based prediction, we identified the known transposable elements (TE) within the *A. corniculatum* genome using RepeatMasker (open-4.0.9) with the Repbase TE library. RepeatProteinMask searches were also conducted using the TE protein database as a query library. For *de novo* prediction, we constructed a *de novo* repeat library of the *A. corniculatum* genome using RepeatModeler to comprehensively conduct, refine and classify consensus models of putative interspersed repeats for the *A. corniculatum* genome (Tarailo-Graovac & Chen, 2009). Furthermore, we performed a *de novo* search for long terminal repeat (LTR) retrotransposons against the *A. corniculatum* genome sequences using Ltr_finder (v.1.0.7) (Xu & Wang, 2007). We also identified tandem repeats using the Tandem Repeat Finder (TRF) package (Benson, 1999) and the noninterspersed repeat sequences, including low-complexity repeats, satellites and simple repeats, using RepeatMasker. Finally, we merged these library files of the two methods and used RepeatMasker to identify the repeat contents.

We conducted annotation of protein-coding genes in the *A. corniculatum* genome using a combination of *de novo*, homology-based and RNA-seq-based prediction. We used Augustus (v.3.3.1) (Stanke *et al.*, 2006) and GlimmerHMM (Majoros *et al.*, 2004) to perform *de novo* gene prediction based on the masked genome. Protein sets were collected and chosen as homology-based evidence from sequenced plants Ericales (*Actinidia chinensis* (Huang *et al.*, 2013), *Camellia sinensis* (Xia *et al.*, 2019), *Diospyros oleifera* (Suo *et al.*, 2020), *Primula veris* (Nowak *et al.*, 2015), *Rhododendron delavayi* (Zhang *et al.*, 2017), *Vaccinium corymbosum* (Colle *et al.*, 2019)) and model plants (*Arabidopsis thaliana* (Lamesch *et al.*, 2012) and *Oryza sativa* (Ouyang *et al.*, 2007)). Clean RNA-seq reads were used as RNA-seq-based evidence; gene structure was estimated using Tophat2 (v.2.1.1) and Cufflinks (v.2.2.1) (Trapnell *et al.*, 2012). Finally, Maker (v.3.00) was used to integrate all evidence to generate nonredundant gene models (Cantarel *et al.*, 2007). Functional annotations were assigned according to the best match of the alignments to the NCBI (NR), SwissProt, TrEMBL, InterPro, the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and Pfam nonredundant protein databases. Transcription factor annotations were performed using iTak (v.1.7) (Zheng *et al.*, 2016).

Genome quality assessment

The quality-filtered reads were used for genome size estimation. First, we generated a 17-mer occurrence distribution of sequencing reads from short libraries using the k-mer method. Then, we estimated genome size using GCE (v.1.0.2) (Liu *et al.*, 2013). Flow cytometry analysis for the measurement of nuclear DNA content was performed in our previous study (Lyu *et al.*, 2018). The genome size of *A. corniculatum* was *c.* 841 Mb (flow cytometry) or 896 Mb (k-mer analysis).

To examine the assembly integrity, the Continuous Long Reads (CLR) subreads were aligned to the assembly using MINIMAP2 (v.2.5) (Li, 2018) and the Illumina short reads were also aligned to the assembly using BWA (Li & Durbin, 2009). We evaluated the completeness of the assembly and gene prediction using Benchmarking Universal Single-Copy Orthologues (BUSCO) v.3.1.0 based on the eudicotyledons_odb10 database (2121 single-copy genes) (Seppey *et al.*, 2019).

Phylogenetic analyses

We used ORTHOFINDER (Emms & Kelly, 2019) to identify orthologous genes from *A. corniculatum* and 10 other angiosperms species: *Primula veris* (Nowak *et al.*, 2015), *Diospyros oleifera* (Suo *et al.*, 2020), *Vaccinium corymbosum* (Colle *et al.*, 2019), *Rhododendron delavayi* (Zhang *et al.*, 2017), *Actinidia chinensis* (Huang *et al.*, 2013), *Camellia sinensis* (Xia *et al.*, 2019), *Erythranthe guttata* (formerly known as *Mimulus guttatus*) (Hellsten *et al.*, 2013), *Vitis vinifera* (Jaillon *et al.*, 2007), *Nelumbo nucifera* (Shi *et al.*, 2020), *Oryza sativa* (Ouyang *et al.*, 2007). We also used the reciprocal BLASTP best hit method for the 11 angiosperm proteins and single-copy genes from the BUSCO dataset. We then merged these two results to identify low-copy genes. We individually aligned low-copy orthologous proteins using MAFFT (Katoh & Standley, 2013) and used the aligned protein sequences to generate codon alignments using PAL2NAL (Suyama *et al.*, 2006). We further trimmed alignments using GBLOCKS 0.91b (Castresana, 2000) and discarded alignments shorter than 150 bp. Based on these alignments, we inferred a phylogenetic tree using RAxML-NG with the GTR+GAMMA+I model, 1000 bootstrap replicates (Kozlov *et al.*, 2019) and *Oryza sativa* as an outgroup (Ouyang *et al.*, 2007). Following its reconstruction, we further dated the tree using MCMCTREE from the PAML (v.4.9j) package with approximate likelihood calculation (Yang, 2007). This method provides a fast and efficient way of analysing large datasets using complex models (Reis & Yang, 2011). The HKY85+G nucleotide substitution model and independent-rates clock model were used in molecular dating. Two reliable fossil calibrations were incorporated (Morris *et al.*, 2018). First, the root node of eudicots and monocots was constrained between 125–247 Myr before present (BP). Second, the common ancestor of eudicots was placed at 119.6–128.63 Myr BP. The MCMC analyses were run for 10 million generations and sampled every 500 generations after a burn-in of 1000 000 iterations. The MCMC analyses were run twice independently to ensure convergence. We used the R package GGTREE to visualise the phylogenetic tree (Yu *et al.*, 2017). The origin and credit for the plant pictures are listed in Table S1.

Gene family evolution in Ericales was examined using CAFE (Mendes *et al.*, 2020). We obtained counts of gene families and genes in seven species of Ericales and *E. guttata* from ORTHOFINDER and removed the large gene families with more than 100 gene copies in one or more species. We identified gene family expansions or contractions only when the gene count change was significant with a P-value < 0.01. According to *A. corniculatum* genome annotations, we assigned these gene families to KEGG pathways and compared these pathways' gene numbers in the expanded families and the whole genome.

Whole-genome duplication (WGD) analyses

To detect the degree of collinearity, we aligned protein sequences between *A. corniculatum* and the more closely related species *P. veris* and within each species using BLASTP (with identity ≥ 30%, e-value < 1E-10, alignment length ≥ 30% of both query and reference sequences). We then identified syntenic blocks containing a minimum of five shared genes using MCSCANX (Wang *et al.*, 2012). Syntenic blocks were visualised by TBTOOLS (Chen *et al.*, 2020) and CIRCOS (Krzywinski *et al.*, 2009). For analyses of the WGD events, we first obtained alignments of all gene pairs as described above. Then we used KAKS_CALCULATOR to calculate synonymous substitution rates (Ks) with the YN substitution model (Wang *et al.*, 2010) and visualised the Ks distributions for paralogous genes within *A. corniculatum* and orthologous genes between *A. corniculatum* and *P. veris*. To identify paralogous genes generated in the lineage-specific WGD on the *A. corniculatum* branch, we filtered all gene pairs except those with Ks in the 0.2–1.1 range for further analysis. We also identified single genes using the duplicate_gene_classifier module from MCSCANX (Wang *et al.*, 2012). With single genes as a control, we performed GO and KEGG enrichment analyses of all filtered paired genes belonging to collinear blocks using BINGO in CYTOSCAPE (v.3.7.2) (Shannon *et al.*, 2003).

To gauge the date of the *A. corniculatum* lineage-specific WGD event, we estimated the WGD event's absolute timing through molecular clock analysis of concatenated gene families, calibrated using species divergence times, similarly to the approach adopted by Clark and Donoghue (Clark & Donoghue, 2017). Based on gene duplicates generated by the WGD, we first identified *A. corniculatum* vs *P. veris* orthogroups that were at exactly 2 : 1 copy number ratio according to collinear relationships identified above. The corresponding orthologues from *R. delavayi* and *V. vinifera* in each orthogroup were identified by searching for the *P. veris* gene's best hit. To further verify a clear signal of the recent WGD event in the selected gene families, we built individual gene trees based on multiple sequence alignments of these four plant species as described above and performed gene tree reconstructions using RAxML-NG (Kozlov *et al.*, 2019). We discarded orthogroups whose topology did not clearly reflect the WGD event signal or was incongruent with the species tree. Of the 552 retained orthogroups, 136 gene families had a clear signal of the lineage-specific WGD event on the *A. corniculatum*

branch. To improve the robustness and precision of estimation, we concatenated 136 multiple sequence alignments with a random combination of each two gene copies in *A. corniculatum*. The nodes were constrained using species divergence times from the phylogenetic tree described above. Molecular clock analysis was conducted on concatenated alignments using the approximate likelihood calculation method in MCMCTREE under the appropriate model (Yang, 2007; Reis & Yang, 2011). We also reconstructed the topology based on our concatenated alignments using RAxML-NG and found it to coincide exactly with the constrained tree. Each analysis was run twice independently to ensure convergence. We also estimated the timings of two recent WGD events in *A. chinensis* using the same approach. The timings of the *V. corymbosum* WGD event, *D. oleifera* WGD event, *E. guttata* WGD event, *N. nucifera* WGD event and γ WGD event have been reported in previous studies (Jiao *et al.*, 2012; Ming *et al.*, 2013; Akagi *et al.*, 2020; Feng C *et al.*, 2020a; Yang *et al.*, 2020).

Transcriptome sequencing and analysis

We collected leaves of mature *Aegiceras corniculatum* plants from upstream to downstream estuarine positions (Sanjiang, SJ; Shanwei, SW; Tashi, TS) in the Dongzhai Harbor National Nature Reserve nursery. Each group of coastal field sites contained three independent biological replicates. We measured salt concentrations of water at the three coastal field sites using a conductivity meter (Seven2Go™ S3; Mettler Toledo). Total RNA was extracted using the modified CTAB method for transcriptome sequencing. RNA-seq libraries were prepared for sequencing on the Illumina NovaSeq 6000 platform to generate 150-bp paired-end reads and yielded *c.* 107.21 Gb of sequence (Table S2).

Clean reads from each sample were aligned to the *A. corniculatum* genome using HISAT2 (v.2.2.0) (Kim *et al.*, 2019). The number of reads mapping to each gene was determined using HTSEQ (v.0.13.5) (Anders *et al.*, 2015). We then identified differentially expressed genes (DEGs) using the R package DESEQ2 with shrinkage estimation for dispersions and fold changes (Love *et al.*, 2014). A *P*-value below 5% and fold change greater than two were set as the significantly differential expression threshold. For the upregulated gene identification in high-salt conditions, it needed to be: (1) upregulated in both high-salt conditions compared with low-salt condition (SJ vs SW and SJ vs TS); and (2) not be downregulated in TS compared with SW. Based on these two criteria, we obtained 1277 genes that were robustly upregulated in high-salt conditions. We reported expression levels as transcripts per million (TPM) values and calculated Pearson's correlation coefficients among all replicates using gene expression values. We also visualised key gene expression profiles using the R package PHEATMAP with *z*-score normalisation in the column direction.

To detect the upregulated genes that might undergo positive selection in *A. corniculatum*, we first identified high-quality orthologues within *A. corniculatum* and its relatives, including *P. veris*, *V. corymbosum*, *R. delavayi*, *A. chinensis*, *C. sinensis*, *D. oleifera* and *E. guttata*. Then we used the branch-site model in the PAML (v.4.9j) package to identify positively selected genes (Yang, 2007), as described in our previous study (Feng X *et al.*, 2020b). After manually removing potential errors in their alignments, we considered the upregulated genes with FDR below 5% as PSGs.

Identification of the key crypto-vivipary gene

We searched for candidate homologues of *DELAY OF GERMINATION1* (*DOG1*) in all seven Ericales genomes against *A. thaliana* genome data (TAIR10, www.arabidopsis.org) using BLASTP. We searched for *A. corniculatum* candidate homologues of both *DOG1* and *DOG1-like* genes and retained all hits with an e-value cutoff of 1E-10. Only the best hits were retained for further analysis from the other six species. We aligned all the homologues together to verify their heme-binding potential. To identify the exclusive orthologue of each *A. corniculatum* candidate in *A. thaliana* as well as other Ericales species, we performed BLASTP searches using *A. corniculatum* as a database and selected the best hit. All alignments mentioned above were implemented using the accurate option (L-INS-i) of MAFFT (Katoh & Standley, 2013). Alignment results were displayed using ALIVIEW (Larsson, 2014). Considering the heme-binding site is essential for the *DOG1* function, we mapped Illumina short reads of *A. corniculatum* to the potential *DOG1* region of two relatives, *P. veris* and *C. sinensis*, and manually checked the alignments. To estimate *DOG1* genetic divergence, we used the Kimura two-parameter method (Kimura, 1980) to calculate the genetic divergence among 21 Asterids and three outgroups (Table S3).

Results

Genome sequencing, assembly and annotation

We sequenced the genome of *Aegiceras corniculatum*, one of the most common mangrove species widespread in the IWP (Fig. S1), by incorporating PACBIO SMRT long reads and Illumina short reads. We used a k-mer analysis (Fig. S2) to estimate genome size at *c.* 896 Mb (Liu *et al.*, 2013), consistent with the *c.* 841 Mb value obtained by flow cytometry in our previous study (Lyu *et al.*, 2018). We generated 139.72 Gb (*c.* 169× coverage) of SMRT long reads and 54.71 Gb (*c.* 66× coverage) of short reads (Table 1). After preliminary assembly, correction and polishing, the final assembly was 827 Mb, consistent with the estimated genome size. The longest scaffold was 13.76 Mb and N50 was 3.87 Mb (Table 1). The GC content of the assembled *A. corniculatum* was 34.37%, similar to closely related species *P. veris* (Nowak *et al.*, 2015).

We examined assembly integrity by aligning CLR subreads to the assembly using MINIMAP2 (Li, 2018) and Illumina short reads using BWA (Li & Durbin, 2009). In total, 88.86% of CLR subreads and 97.54% of Illumina reads were successfully mapped back to our genome. Based on short-read mapping, we estimated heterozygosity at *c.* 2.044 sites per kb. We also evaluated the completeness of our genome assembly using BUSCO v.3.1.0 with the eudicotyledons_odb10 lineage dataset (Seppey *et al.*, 2019).

Table 1 *Aegiceras corniculatum* genome assembly and gene annotation statistics.

Genome features	
SMRT long reads	139.72 Gb
Illumina paired-end reads	54.71 Gb
Estimated genome size	896 Mb
Assembled genome size	826.70 Mb
Scaffold number	1494
N50 length	3.87 Mb
N50 count	64
Longest scaffold	13.76 Mb
Number of scaffolds (≥1 kb)	1488
GC content	34.37%
Number of genes	32092

The BUSCO analysis showed that, at the genomic level, 91.8% of the 2121 expected plant genes were identified as complete, while 6.1% were missing from the assembled *A. corniculatum* genome (Table S4). These results indicated that our genome assembly was of high quality.

We estimated that 72.10% of the *A. corniculatum* reference genome consisted of repetitive sequences. These are predominantly TE families, comprising 560.62 Mb (67.81%) of the genome (Table S5). With repetitive sequences masked, we predicted 32 092 protein-coding genes (Table 1; Fig. S3). We were able to functionally annotate 29 207 of them by comparison with public genomic resources from other plants (Table S6). We also predicted 816 tRNAs, 963 rRNAs, 157 snRNAs and 56 miRNAs.

Phylogenetic position

To determine the position of *A. corniculatum* within Ericales, we reconstructed the phylogeny of 11 angiosperm species with available genome sequences. After searching, aligning and filtering, we identified 1480 orthologous low-copy genes. Based on these orthologues, we inferred a phylogenetic tree with RAxML-NG using the GTR+GAMMA+I model (Kozlov *et al.*, 2019) and *Oryza sativa* (Ouyang *et al.*, 2007) as the outgroup (Fig. S4). The phylogenetic relationships among and within the main clades were consistent with previous studies (Chase *et al.*, 2016; Zhang *et al.*, 2017; Yang *et al.*, 2020). We estimated that *Aegiceras* diverged from *Primula c.* 42.59 Ma, while Primulaceae diverged from the common ancestor of Ebenaceae, Ericaceae, Actinidiaceae and Theaceae *c.* 73.81 Ma (Fig. 1).

Gene family analysis

We compared the complexity of gene families between *A. corniculatum* and other species (*P. veris* (Nowak *et al.*, 2015), *D. oleifera* (Suo *et al.*, 2020), *V. corymbosum* (Colle *et al.*, 2019), *R. delavayi* (Zhang *et al.*, 2017), *A. chinensis* (Huang *et al.*, 2013), *C. sinensis* (Xia *et al.*, 2019) and *E. guttata* (Hellsten *et al.*, 2013)). We identified 8291 gene families common among these species. Of these, 2663 showed signs of expansion in *A. corniculatum*, while another 3096 lost members in this lineage. However, only 37 gene families showed statistically significant expansion using CAFE (Mendes *et al.*, 2020) in *A. corniculatum* and one gene family that contracted since the MRCA of *A. corniculatum* and *P. veris* (Table S7). Gene families encoding proteins involved in ATP-binding cassette transport, oxidative-phosphorylation and photosynthesis have most obviously expanded (Table S8). The gene families related to natural antioxidant biosynthesis and plant defence also expanded in *A. corniculatum* (Table S9).

Whole-genome duplication

Whole-genome duplication (WGD) has been widely observed in plants and animals (Dehal & Boore, 2005; Van de Peer *et al.*, 2017; Clark & Donoghue, 2018). Polyploid plants may be able to survive and even thrive in extreme environments (Wu *et al.*, 2020; He *et al.*, 2020; Van de Peer *et al.*, 2020). Therefore, we wondered whether *Aegiceras* had undergone any recent WGDs. We scanned the genomes of *A. corniculatum* and a closely related species, *P. veris*, using BLASTP and MCSCANX. We found that the *A. corniculatum* genome contains 254 syntenic block pairs with at least five shared genes, comprising 6154 (19.18%) protein-coding genes (Fig. S5). This is likely to be an underestimate of the full extent of genomic duplication, as lost duplicates do not count in this measure. If we include all genes residing within a block delineated by five or more duplicated loci, these blocks cover 18 324 (57.10%) genes. By contrast, the *P. veris* genome has only 12 syntenic blocks with 195 (1.07%) genes in them. We also identified 578 such block pairs between the two species. The excess of within-species syntenic block number indicates a possible WGD event within *A. corniculatum* (Fig. 2a). To time this event, we calculated synonymous substitution rates (Ks) between the paralogous genes within the *A. corniculatum* genome. As controls, we looked at Ks distribution between the two species. The mode of the Ks distribution between paralogues within *A. corniculatum* is at a smaller value than for *A. corniculatum* vs *P. veris* (Fig. 2b), reflecting higher similarity and therefore indicating that the WGD happened after the divergence between *Aegiceras* and *Primula*. To further estimate the absolute timing of this *A. corniculatum* lineage-specific WGD event, we performed a molecular clock analysis of concatenated gene families (strictly filtered, see Materials and Methods), calibrated using between-species divergence time. The event was placed at *c.* 35 Ma (Fig. 1).

Genes acquired during a large-scale duplication event are typically rapidly lost. Retained duplicates are potentially enriched for genes whose increased expression or acquired functions aid survival in the newly acquired habitat (Freeling, 2009; Van de Peer *et al.*, 2017). With single-copy genes as a control, we performed GO enrichment and pathway enrichment analyses on the genes retained after the *Aegiceras*-specific WGD event. We found that most of the genes were involved in response to stimulus, protein phosphorylation, inorganic ion transmembrane transport, ATP binding, protein kinase activity, transcription regulator activity, calcium ion binding and mitochondrial proton-transporting ATP synthase complex GO categories (Fig. 2c). The NOD-like-receptor signalling, MAPK signalling, calcium signalling and

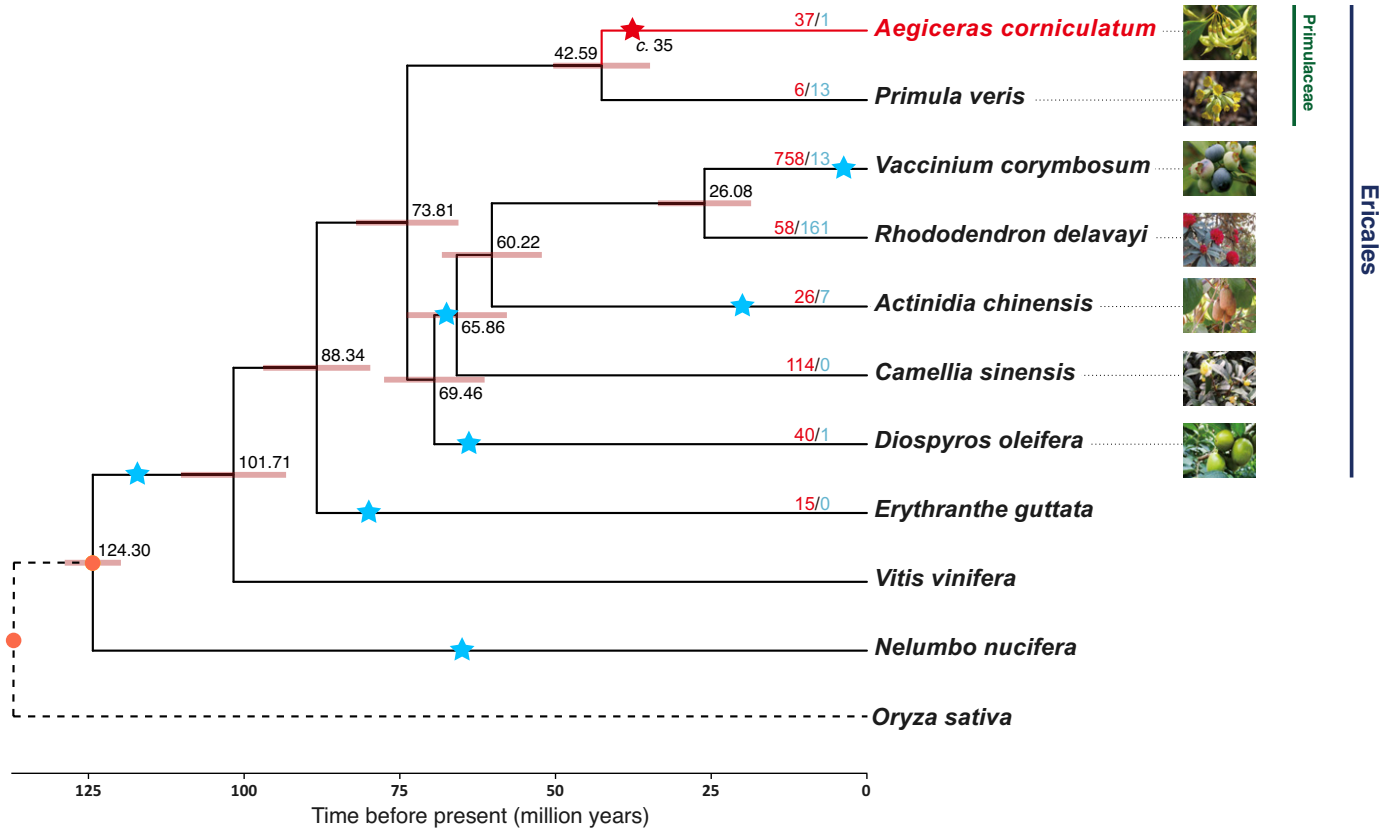

Fig. 1 Phylogenetic tree of 11 angiosperm species, including *A. corniculatum* and relatives. The estimated divergence time is shown beside each node. Pink node bars are 95% confidence intervals. Red and blue numbers indicate significantly expanded and contracted gene families. The red nodes indicate two fossil calibration nodes. The dashed line denotes the outgroup *O. sativa*. The asterisks represent the phylogenetic position and estimated date of WGD events. The timings of the *A. corniculatum* WGD event and two recent WGD events in *A. chinensis* are estimated in this study, and other reported WGD timings are superimposed on the phylogenetic tree. The credits for morphology pictures are listed in Supporting information Table S1.

plant hormone signal transduction pathways were also over-represented (Table S10). The preferential retentions of transcription regulation, signal transduction and energy metabolism pertained to adaptation to intertidal zones.

High-salt adaptation-supportive mechanisms

The most extreme challenge for species living in intertidal zones is the unstable and high salinity due to tidal fluctuations. *A. corniculatum* has evolved salt glands on leaves to enhance its salt tolerance. Among the genes retained after the recent WGD, the calcium-activated 14-3-3 protein-coding gene is particularly interesting as it is a molecular switch in salt stress, decoding a calcium signal to enhance plant salt tolerance (Yang *et al.*, 2019). After a strict alignment with the Pfam and NR databases, and the *A. thaliana* annotation, we identified 14 14-3-3 protein-coding genes in *A. corniculatum*, more than the nine in *P. veris* (Table S11). We also identified 19 H^+-ATPase coding genes in *A. corniculatum*, more than in its relatives (Table S12). After the recent WGD event, 12 duplicates of 14-3-3 protein-coding genes and four duplicates of H^+-ATPase coding genes were retained. We noted that many retentions were annotated with functions involving calcium signal-activated SOS pathway, in particular *J3* and calcineurin-related genes (Fig. 3a). This result confirmed the GO enrichment and pathway enrichment analyses and showed preferential recent WGD retentions of signal transduction and energy metabolism during salt adaptation.

To further investigate the mechanisms underlying salt tolerance in *A. corniculatum*, we sequenced leaf transcriptomes at three coastal field sites (Fig. 3c; Table S13), including one low-salt (SJ, 5.15‰ water salinity) and two high-salt conditions (SW, 21.42‰; TS, 26.25‰). Using the HISAT2–HTSEQ–DESEQ2 workflow (Love *et al.*, 2014; Anders *et al.*, 2015; Kim *et al.*, 2019), we obtained expression profiles and identified DEGs (Fig. S6; Table S14). We found 1277 genes that were significantly upregulated in the high-salt (SW and TS) vs low-salt (SJ) environment, but not downregulated across salinity conditions (see Materials and Methods for details). Most genes involved in the maintenance and regulation of cellular environmental homeostasis were differentially regulated. In total, 223 of these genes were retained after the lineage-specific WGD event, including 29 duplicated gene pairs and 165 retentions with only one copy upregulated under high salt, enriched in transcriptional regulation and signal transduction. As in our comparative genomic analysis, H^+-ATPase protein-coding genes have more copies in *A. corniculatum* than in its relatives. Expression of one of these, *Aco020046* (*HA5*), increased in high-salt environments.

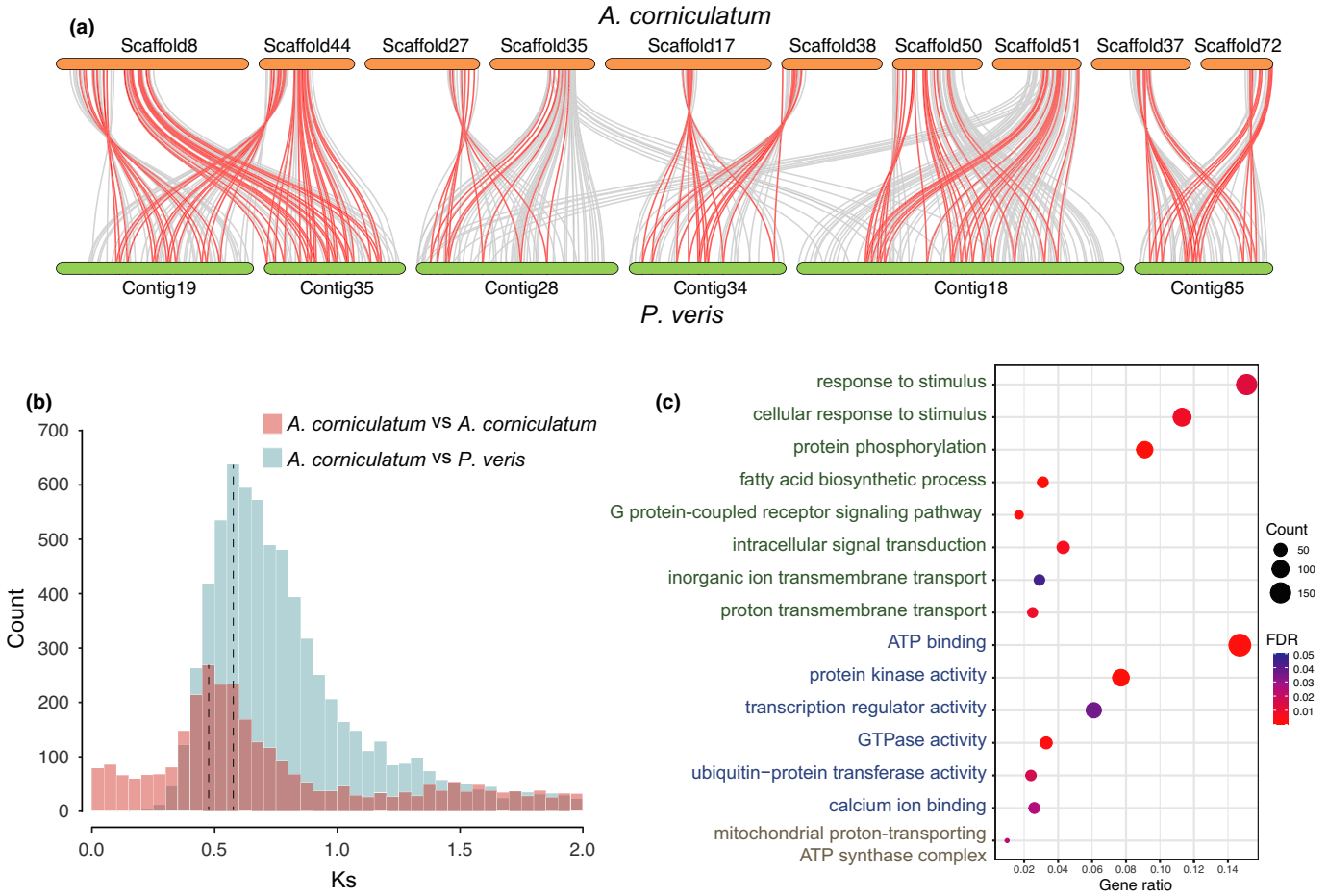

Fig. 2 The *A. corniculatum* lineage-specific whole-genome duplication (WGD) event. (a) Diagram of syntenic blocks between genomic regions from *A. corniculatum* and the closely related species *P. veris*. Red links represent collinear gene pairs with a 1 : 2 ratio in the two species. (b) Ks distribution between paralogous genes within the same species and orthologous genes between the two species. Dashed lines denote the WGD event (left) and speciation event (right). (c) Gene Ontology enrichment among retained gene duplicates compared with single genes. The size and colour of the bubbles represent gene number and FDR value calibrated using the Benjamini–Hochberg method. A bubble colour scale closer to red indicates a smaller FDR value,. The GO categories are shown on the *y*-axis. The green, blue and grey labels indicate Biological Process, Molecular Function and Cellular Component, respectively.

This protein can help to establish the proton motive force across the plasma membrane and participate in salt secretion, therefore maintaining a low Na^+ concentration inside cells. Other genes upregulated by high salt were ion carriers or channels, or were involved in vesicle transport, trichome development, biosynthesis of callose (the most important component of plasmodesmata), and biosynthesis of cuticles (Fig. 3d; Table S15). Increased expression of these genes may promote salt transport efficiency from the mesophyll to the salt gland and subsequent secretion (Fig. 3a). We also found that upregulated transcripts over-represented among antioxidant biosynthesis pathway genes. The glutathione, flavonoid, superoxide dismutase (SOD) and peroxidase (POD) pathways can contribute to scavenging reactive oxygen species (ROS) against secondary oxidation stress caused by hypersaline stress (Fig. 3b,d; Table S16). Several genes coded for transcription factors, such as WRKY, bHLH, NAC and AP2/ERF (Fig. 3b), which play roles in the regulation of flavonoid biosynthesis and glutathione metabolism pathways (Morishita *et al.*, 2009; Mizoi *et al.*, 2012; Xu *et al.*, 2015; Jiang *et al.*, 2017). In addition, we found that 12 upregulated genes had probably undergone positive selection in *A. corniculatum* (Table S17). These were involved in calcium signalling transduction, cytoskeleton component, cellular environmental homeostasis maintenance and relevant to salinity resistance.

Identification of a pivotal gene function loss accounting for crypto-vivipary

To explore genetic mechanisms underlying the emergence of crypto-vivipary in *A. corniculatum*, we focused on homologues of *DELAY OF GERMINATION1* (*DOG1*), a pivotal regulator controlling seed dormancy (Nishimura *et al.*, 2018; Nonogaki, 2019). We searched for putative homologues of *AtDOG1* in each Ericales genome and selected the best hits. Despite the broad sequence diversity of *DOG1* (Table S18), we discovered a conspicuous amino acid divergence that appeared exclusively in one *A. corniculatum* homologue Aco004515 within a relatively conserved region at position

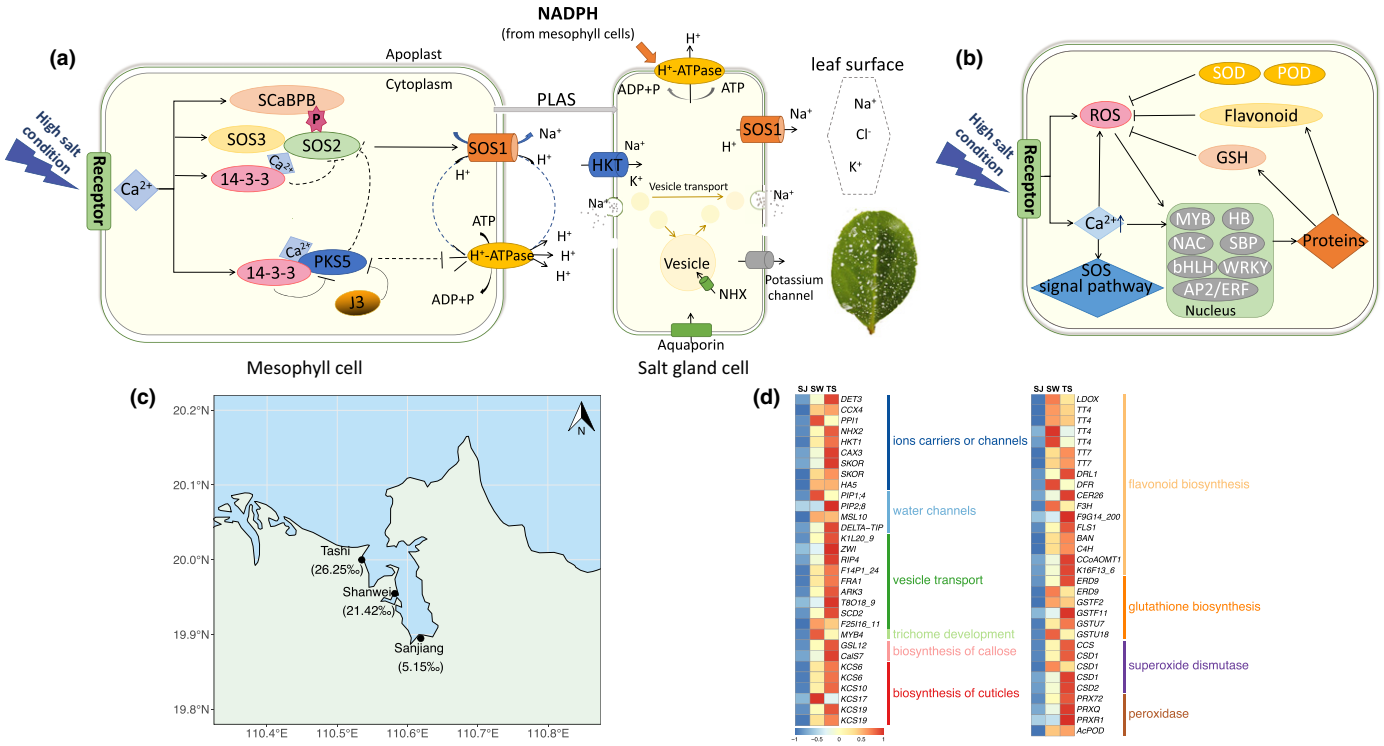

Fig. 3 High-salt adaptation of *A. corniculatum*. (a) The major pathway of Na^+ transport and salt secretion in *A. corniculatum*. The 14-3-3 protein-coding gene, decoding a calcium signal to enhance plant salt tolerance, can activate the SOS pathway under high salinity conditions. Na^+ is transported from the mesophyll to the salt gland and then secreted onto the leaf surface. Based on Yuan *et al*. (2016) and Yang *et al*. (2019). (b) Natural antioxidant biosynthesis can contribute to scavenging ROS against high salinity. (c) Sampling locations. The salt concentrations of water at the three coastal field sites are shown in parentheses. (d) Expression heatmap of upregulated genes under high-salt conditions involved in salt homeostasis and ROS scavenging. The scale on the bottom left represents TPM values with *z*-score normalisation.

633 of the multisequence alignment. All other genes had a histidine residue at this position, while this was replaced by glutamine in Aco004515 of *A. corniculatum* (Fig. S8a, see later). This residue, His^{245} of AtDOG1, is critical for heme-binding site formation. Together with His^{249}, His^{245} serves as an axial ligand for the heme (Fig. 4a). A transgenic experiment demonstrated that heme binding occurs via His^{245} and His^{249} in Arabidopsis (Nishimura *et al*., 2018).

We inferred that the incompleteness of the heme-binding site caused by missing one of the two histidine residues might result in a functional deficiency of *DOG1* homologues in *A. corniculatum* (Fig. 4b). Considering the genetic divergence of *DOG1* among Asterids plant species (Fig. S7; Tables S3,S18) and the abundance of *DOG1-like* homologues in *A. corniculatum*, we wondered whether there is any potential compensation achieved by other *DOG1* or *DOG1-like* homologues with a complete heme-binding site. Therefore, we searched for all potential candidate *DOG1* genes (see Materials and Methods). We found that all candidates appeared to have lost critical histidine residues (Fig. 4c). For each candidate, we further searched for the best hit in *A. thaliana* as well as other Ericales genomes (Fig. S8b–f). To exclude the possibility of assembly failure, we mapped more accurate Illumina short reads we generated for *A. corniculatum* to the *DOG1* region of two closely related species, *P. veris* and *C. sinensis*. These alignments showed the same loss of critical His amino acids. These results indicated that no functional heme-binding *DOG1* homologue exists in *A. corniculatum*.

Discussion

As a pioneer mangrove species in the Primulaceae family, *A. corniculatum* has a series of specialised adaptive traits, including salt secretion and crypto-vivipary. A combination of Ks-based, synteny and molecular clock analyses shows that *A. corniculatum* experienced a recent round of WGD since the well known paleo-hexaploidisation (γ WGD) event shared by core eudicots (Jiao *et al*., 2012). Genes preferentially retained after this duplication are disproportionately involved in stimulus response, ion transport, signal transduction and energy metabolism. The evidence presented here suggested that the lineage-specific WGD event might have contributed to the adaptation of *A. corniculatum* to the fluctuating high salinity of tropical intertidal environments.

In general, salt adaptation is both a long-term and dynamic process and involves many morphological, physiological, cellular and molecular processes (Munns & Tester, 2008; Feng X *et al*., 2020b). In *Populus euphratica*, numerous genes involved in ion transport, transcription factor activity, oxidoreductase activity and hormone biosynthesis are differentially expressed under salt stress (Ma *et al*., 2013; Zhang *et al*., 2014). In *Zostera marina*, the key genes, Na^+ and K^+ antiporters, are displayed as a typical

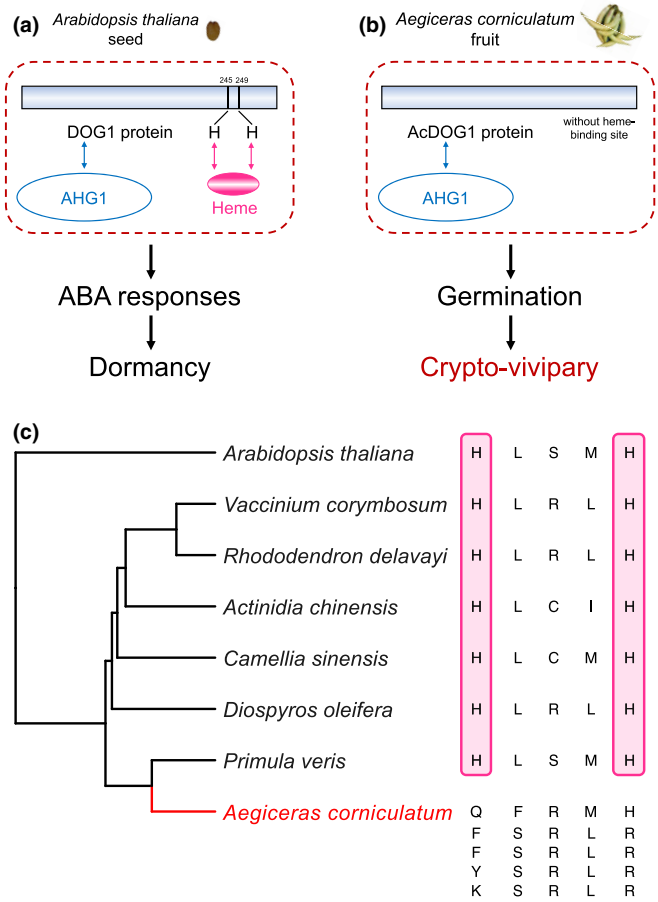

Fig. 4 All DOG1 homologues lost the ability to bind heme in *A. corniculatum* leading to functional deficiency. (a) Mechanism of seed dormancy induced by AtDOG1 with the heme-binding site. Based on Nishimura *et al*. (2018) and Nonogaki (2019). (b) Loss of the heme-binding site in the homologous proteins prompts seed germination, probably accounting for the specialised crypto-vivipary trait. (c) Amino acid alignment of the heme-binding site among AtDOG1 proteins, putative homologues in each Ericales genome. The red frame highlights the two critical histidine residues of the heme-binding site.

component and strongly expressed in vegetative tissue (Muramatsu *et al.*, 2002; Olsen *et al.*, 2016). Mangrove trees can induce specific salt-responsive genes and transcription factors to facilitate adaptation to intertidal environments (Feng X *et al.*, 2020b). The SOS pathway is a well defined signalling pathway for controlling ion homeostasis at the cellular and tissue level (Ji *et al.*, 2013). The 14-3-3 protein decoding salt-induced calcium signals can activate SOS1 (plasma membrane Na^+/H^+ antiporter) and PM H^+-ATPase to adapt to salt stress (Yang *et al.*, 2019). Our comparative genomic analysis revealed that 14-3-3 and H^+-ATPase protein-coding genes have more copies in *A. corniculatum* than in its relatives (*P. veris*, *E. guttata* and *V. vinifera*). Salt secretion by glands is a highly specialised trait in mangroves to enhance their salt tolerance (Shi *et al.*, 2005). This mechanism involves an increase in H^+-ATPase expression as salt secretion accelerates in high-salt environments (Chen *et al.*, 2010). Salt secretion places high energy demands on the plant. Mesophyll cells can provide photoassimilates and NADPH (the sources of energy for salt excretion) to the salt gland cells by plasmodesmata (Yuan *et al.*, 2016).

The biggest challenge for mangrove trees surviving in the extreme and harsh intertidal environment is the unstable and high salinity due to tidal fluctuations. Our comparative genomic and transcriptomic analyses suggested that the maintenance of cellular environmental homeostasis is an important adaptive process in *A. corniculatum*. Salt ions are transported from mesophyll cells into the salt gland and then secreted out (Fig. 3a). Hypersaline stress always leads to secondary oxidation stress. Expansion and upregulation of antioxidant biosynthesis-related genes can contribute to scavenging ROS against hypersaline stress (Fig. 3b). The functional enrichment of genes retained after WGD also supports homeostasis maintenance as an important adaptation driver.

While the high-salt supportive mechanisms enable *A. corniculatum* as well as other mangroves to thrive in intertidal zones, another main obstacle is to reproduce and colonise in this harsh and unstable environment, as mangrove seeds are exposed to many additional unfavourable conditions, including high temperature, waterlogging, hypoxia and other stresses. Unlike regular seed development, vivipary or crypto-vivipary in mangrove seeds can decrease dormancy and promote germination when still attached to the mother plant. It is now well established that *DOG1* is a key regulator controlling seed maturation, dormancy and germination in plants (Nakabayashi *et al.*, 2012; Nonogaki, 2019). *DOG1* belongs to a small gene family consisting of *DOG1* and four additional *DOG1-like* genes. They are found in many species such as rice (Sugimoto *et al.*, 2010), wheat, barley, sorghum (Ashikawa *et al.*, 2013) and Brassicaceae (Graeber *et al.*, 2010) including *A. thaliana* (Bentsink *et al.*, 2006). Since *DOG1* was first identified and cloned in *A. thaliana* as a major quantitative trait locus for seed dormancy (Bentsink *et al.*, 2006), most mechanisms of seed dormancy have been elucidated in this species (Nonogaki, 2019). Recent genomic analyses in mangrove trees has suggested that amino acid substitution and positive selection of essential seed development genes, expression changes of key plant hormone metabolism and embryonic genes, reduction of proanthocyanidin and storage protein production, and *DOG1* loss together contribute to vivipary in the Rhizophoraceae family (Xu *et al.*, 2017; Qiao *et al.*, 2020). In the crypto-viviparous mangrove *A. corniculatum*, embryos only break out of the seed coat before dehiscence. Considering the genetic divergence of *DOG1* among Asterids plant species (Tables S3,S18) and the wide distribution of functional *DOG1-like* genes in cereals (Ashikawa *et al.*, 2013), we searched all potential *DOG1* or *DOG1-like* homologues in *A. corniculatum* and its relatives. The complete heme-binding site found in other species appears to have been lost in *A. corniculatum* (Fig. 4c). Previous studies have revealed that heme binding at His^{245} and His^{249} of DOG1 is critical to its seed dormancy function (Nishimura *et al.*, 2018). Therefore, losing the complete heme-binding site promotes seed germination, probably enabling crypto-vivipary (Fig. 4b), protecting propagules from stresses during early development. This strategy can enhance *A. corniculatum*'s fecundity in an unstable and extreme environment.

In summary, we constructed a high-quality genome assembly for *Aegiceras corniculatum* by combining SMRT long reads with highly accurate short reads. This provides new insights into the evolution of two specialised adaptive traits (salt secretion and crypto-vivipary) in mangroves. We deduced that the maintenance of cellular environmental homeostasis is an important adaptive process. Our study also will be helpful for genetic, genomic and evolutionary studies in both *Aegiceras* and other plants inhabiting extreme environments.

Acknowledgements

We thank Anthony J. Greenberg for manuscript modifications. The project was supported by the National Natural Science Foundation of China (31971540 and 31830005); the National Key Research and Development Plan (2017FY100705); the Guangdong Basic and Applied Basic Research Foundation (2019A1515010752); the Science and Technology Project of Guangzhou (202102020483); and the Chang Hungta Science Foundation of Sun Yat-sen University.

Author contributions

ZH and S. Shi conceived the study; XF, GL, SX, WW, QC, S. Shao, NW and ZH analysed the data; XF, ML and CZ collected materials; XF, GL and ZH wrote the manuscript; all authors read and approved the final manuscript.

ORCID

Xiao Feng https://orcid.org/0000-0001-8919-6000

Data availability

The raw genomic Illumina reads, PacBio reads and RNA-seq reads reported in this paper have been deposited in the Genome Sequence Archive (GSA, https://bigd.big.ac.cn/gsa) in the National Genomics Data Center (CNCB-NGDC Members & Partners, 2021), Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession number CRA003915 with BioProject ID PRJCA004487. The genome assembly sequences have been deposited in the Genome Warehouse (GWH, https://bigd.big.ac.cn/gwh) in National Genomics Data Center under accession number GWHBAOL00000000 with BioProject ID PRJCA004487 and BioSample ID SAMC339266.

References

Akagi T, Shirasawa K, Nagasaki H, Hirakawa H, Tao R, Comai L, Henry IM. 2020. The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. *PLoS Genetics* **16**: e1008566.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Ashikawa I, Abe F, Nakamura S. 2013. *DOG1*-like genes in cereals: investigation of their function by means of ectopic expression in *Arabidopsis*. *Plant Science* **208**: 1–9.

Ball MC. 1988. Salinity tolerance in the mangroves *Aegiceras corniculatum* and *Avicennia marina*. I. water use in relation to growth, carbon partitioning, and salt balance. *Functional Plant Biology* **15**: 447–464.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573–580.

Bentsink L, Jowett J, Hanhart CJ, Koornneef M. 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **103**: 17042–17047.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M. 2007. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–196.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF *et al.* 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* **13**: 1194–1202.

Chen J, Xiao Q, Wu F, Dong X, He J, Pei Z, Zheng H. 2010. Nitric oxide enhances salt secretion and Na^+ sequestration in a mangrove plant, *Avicennia marina*, through increasing the expression of H^+-ATPase and Na^+/H^+ antiporter under high salinity. *Tree Physiology* **30**: 1570–1585.

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A *et al.* 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* **13**: 1050–1054.

Clark JW, Donoghue PCJ. 2017. Constraining the timing of whole genome duplication in plant evolutionary history. *Proceedings of the Royal Society B: Biological Sciences* **284**: 20170912.

Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. *Trends in Plant Science* **23**: 933–945.

CNCB-NGDC Members and Partners. 2021. Database resources of the national genomics data center, china national center for bioinformation in 2021. *Nucleic Acids Research* **49**: D18–D28.

Colle M, Leisner CP, Wai CM, Ou S, Bird KA, Wang J, Wisecaver JH, Yocca AE, Alger EI, Tang H *et al.* 2019. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *GigaScience* **8**: giz012.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* **3**: e314.

Reis RM, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**: 2161–2172.

Elmqvist T, Cox PA. 1996. The evolution of vivipary in flowering plants. *Oikos* 77: 3–9.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Feng C, Wang J, Wu L, Kong H, Yang L, Feng C, Wang K, Rausher M, Kang M. 2020a. The genome of a cave plant, *Primulina huaijiensis*, provides insights into adaptation to limestone karst habitats. *New Phytologist* **227**: 1249–1263.

Feng X, Xu S, Li J, Yang Y, Chen Q, Lyu H, Zhong C, He Z, Shi S. 2020b. Molecular adaptation to salinity fluctuation in tropical intertidal environments of a mangrove tree *Sonneratia alba*. *BMC Plant Biology* **20**: 178.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* **60**: 433–453.

Ge XJ, Sun M. 1999. Reproductive biology and genetic diversity of a cryptoviviparous mangrove *Aegiceras corniculatum* (Myrsinaceae) using allozyme and intersimple sequence repeat (ISSR) analysis. *Molecular Ecology* **8**: 2061–2069.

Giri C, Ochieng E, Tieszen LL, Zhu Z, Singh A, Loveland T, Masek J, Duke N. 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography* **20**: 154–159.

Graeber K, Linkies A, Müller K, Wunchova A, Rott A, Leubner-Metzger G. 2010. Cross-species approaches to seed dormancy and germination: conservation and biodiversity of ABA-regulated mechanisms and the Brassicaceae *DOG1* genes. *Plant Molecular Biology* **73**: 67–87.

He Z, Li X, Yang M, Wang X, Zhong C, Duke NC, Wu C-I, Shi S. 2019. Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *National Science Review* **6**: 275–288.

He Z, Xu S, Zhang Z, Guo W, Lyu H, Zhong C, Boufford DE, Duke NC, Shi S. 2020. Convergent adaptation of the genomes of woody plants at the land-sea interface. *National Science Review* **7**: 978–993.

Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences, USA* **110**: 19478–19482.

Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M *et al.* 2013. Draft genome of the kiwifruit *Actinidia chinensis*. *Nature Communications* **4**: 2640.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

Ji H, Pardo JM, Batelli G, Van Oosten MJ, Bressan RA, Li X. 2013. The Salt Overly Sensitive (SOS) pathway: established and emerging roles. *Molecular Plant* **6**: 275–286.

Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. 2017. WRKY transcription factors in plant responses to stresses. *Journal of Integrative Plant Biology* **59**: 86–101.

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ *et al.* 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biology* **13**: R3.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**: 907–915.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–120.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**: 722–736.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* **19**: 1639–1645.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M *et al.* 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–D1210.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**: 3276–3278.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

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

Liang S, Zhou R, Dong S, Shi S. 2008. Adaptation to salinity in mangroves: implication on the evolution of salt-tolerance. *Science Bulletin* **53**: 1708–1715.

Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv:* 1308.2012v2.

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 H, He Z, Wu C-I, Shi S. 2018. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *New Phytologist* **217**: 428–438.

Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J *et al.* 2013. Genomic insights into salt adaptation in a desert poplar. *Nature Communications* **4**: 2797.

Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**: 2878–2879.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**: 5516–5518.

Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li L-T, Zhang Q, Kim M-J, Schatz MC, Campbell M *et al.* 2013. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biology* **14**: R41.

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta* **1819**: 86–96.

Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. 2009. Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant and Cell Physiology* **50**: 2210–2222.

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **115**: E2274–E2283.

Munns R, Tester M. 2008. Mechanisms of salinity tolerance. *Annual Review of Plant Biology* **59**: 651–681.

Muramatsu Y, Harada A, Ohwaki Y, Kasahara Y, Takagi S, Fukuhara T. 2002. Salt-tolerant ATPase activity in the plasma membrane of the marine angiosperm *Zostera marina* L. *Plant and Cell Physiology* **43**: 1137–1145.

Nakabayashi K, Bartsch M, Xiang Y, Miatton E, Pellengahr S, Yano R, Seo M, Soppe WJJ. 2012. The time required for dormancy release in *Arabidopsis* is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds. *Plant Cell* **24**: 2826–2838.

Nishimura N, Tsuchiya W, Moresco JJ, Hayashi Y, Satoh K, Kaiwa N, Irisa T, Kinoshita T, Schroeder JI, Yates JR *et al.* 2018. Control of seed dormancy and germination by DOG1-AHG1 PP2C phosphatase complex via binding to heme. *Nature Communications* **9**: 2132.

Nonogaki H. 2019. Seed germination and dormancy: the classic story, new puzzles, and evolution. *Journal of Integrative Plant Biology* **61**: 541–563.

Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biology* **16**: 12.

Olsen JL, Rouzé P, Verhelst B, Lin Y-C, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F *et al.* 2016. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**: 331–335.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek Rl, Lee Y, Zheng L *et al.* 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* **35**: D883–D887.

Parida AK, Jha B. 2010. Salt tolerance mechanisms in mangroves: a review. *Trees* **24**: 199–217.

Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2020. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**: 11–26.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**: 411–424.

Qiao H, Zhou X, Su W, Zhao X, Jin P, He S, Hu W, Fu M, Yu D, Hao S *et al.* 2020. The genomic and transcriptomic foundations of viviparous seed development in mangroves. *bioRxiv*. doi: 10.1101/2020.10.19.346163.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 460.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology* **1962**: 227–245.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498–2504.

Shi S, Huang Y, Zeng K, Tan F, He H, Huang J, Fu Y. 2005. Molecular phylogenetic analysis of mangroves: independent evolutionary origins of vivipary and salt secretion. *Molecular Phylogenetics and Evolution* **34**: 159–166.

Shi T, Rahmani RS, Gugger PF, Wang M, Li H, Zhang Y, Li Z, Wang Q, Van de Peer Y, Marchal K *et al.* 2020. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Molecular Biology and Evolution* **37**: 2394–2413.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435–W439.

Sugimoto K, Takeuchi Y, Ebana K, Miyao A, Hirochika H, Hara N, Ishiyama K, Kobayashi M, Ban Y, Hattori T *et al.* 2010. Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proceedings of the National Academy of Sciences, USA* **107**: 5792–5797.

Suo Y, Sun P, Cheng H, Han W, Diao S, Li H, Mai Y, Zhao X, Li F, Fu J. 2020. A high-quality chromosomal genome assembly of *Diospyros oleifera* Cheng. *GigaScience* **9**: giz164.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609–W612.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **25**: 1–14.

Tomlinson PB. 2016. *The botany of mangroves*, 2nd edn. Cambridge, UK: Cambridge University Press.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**: 562–578.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Research* **27**: 737–746.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: e112963.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics* **8**: 77–80.

Wang Y, Tang H, DeBarry Jd, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H *et al.* 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**: e49.

Wu S, Han B, Jiao Y. 2020. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Molecular Plant* **13**: 59–71.

Xia E-H, Li F-D, Tong W, Li P-H, Wu Q, Zhao H-J, Ge R-H, Li R-P, Li Y-Y, Zhang Z-Z *et al.* 2019. Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnology Journal* **17**: 1938–1953.

Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, Han Y, Luo F, Xie Z. 2017. MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nature Methods* **14**: 1072–1074.

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 W, Dubos C, Lepiniec L. 2015. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science* **20**: 176–185.

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 F-S, Nie S, Liu H, Shi T-L, Tian X-C, Zhou S-S, Bao Y-T, Jia K-H, Guo J-F, Zhao W *et al.* 2020. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nature Communications* **11**: 5269.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang Z, Wang C, Xue Y, Liu X, Chen S, Song C, Yang Y, Guo Y. 2019. Calcium-activated 14-3-3 proteins as a molecular switch in salt stress tolerance. *Nature Communications* **10**: 1199.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. GGTREE: an package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**: 28–36.

Yuan F, Leng B, Wang B. 2016. Progress in studying salt secretion from the salt glands in recretohalophytes: how do plants secrete salt? *Frontiers in Plant Science* **7**: 977.

Zhang J, Feng J, Lu J, Yang Y, Zhang X, Wan D, Liu J. 2014. Transcriptome differences between two sister desert poplar species under salt stress. *BMC Genomics* **15**: 337.

Zhang Lu, Xu P, Cai Y, Ma L, Li S, Li S, Xie W, Song J, Peng L, Yan H *et al.* 2017. The draft genome assembly of *Rhododendron delavayi* Franch. var. delavayi. *GigaScience* **6**: gix076.

Zheng W, Wang W, Lin P. 1999. Dynamics of element contents during the development of hypocotyles and leaves of certain mangrove species. *Journal of Experimental Marine Biology and Ecology* **233**: 247–257.

Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ *et al.* 2016. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant* **9**: 1667–1670.

Supporting Information

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

Fig. S1 Morphology and distribution of *A. corniculatum*.

Fig. S2 The k-mer distribution of the *A. corniculatum* genome.

Fig. S3 Summary of predicted protein-coding genes' features.

Fig. S4 Maximum likelihood tree of 11 angiosperm species, including *A. corniculatum* and relatives.

Fig. S5 Diagram of syntenic blocks in *A. corniculatum* genome.

Fig. S6 Gene expression correlations among all samples.

Fig. S7 Phylogeny of 24 plant species (21 Asterids and three outgroups).

Fig. S8 Amino acid alignment of the heme-binding site among AtDOG1 proteins, putative homologues in each Ericales genome.

Table S1 Plant image credits.

Table S2 RNA-seq data information.

Table S3 Summary information of 21 Asterids and three outgroups.

Table S4 Genome assembly and annotation completeness.

Table S5 Transposable elements in the *A. corniculatum* genome.

Table S6 Functional annotation of predicted genes in the *A. corniculatum* genome.

Table S7 Gene family expansion and contraction summary.

Table S8 KEGG pathways over-represented among genes in significantly expanded gene families.

Table S9 Functional annotation of significantly expanded gene families in *A. corniculatum*.

Table S10 Summary of KEGG pathways over-represented among genes in WGD retentions.

Table S11 Summary of *14-3-3* genes.

Table S12 Summary of H^+-ATPase protein-coding genes.

Table S13 Salt concentration measurements of water in three coastal field sites.

Table S14 Numbers of differentially expressed genes.

Table S15 Detailed annotation of upregulated genes under high salt related to salt transport and secretion.

Table S16 Detailed annotation of ROS scavenging genes upregulated under high salt.

Table S17 Detailed annotation of upregulated positively selected genes under high salt.

Table S18 Summary of K2P genetic divergences among 24 plants (21 Asterids and 3 outgroups).

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.