PNAS

Supporting Information for

Scattered differentiation of unlinked loci across the genome underlines ecological divergence of the selfing grass *Brachypodium stacei*

Wenjie Mu[a,b,c,1], Kexin Li[a,1], Yongzhi Yang[a,1], Adina Breiman[d], Jiao Yang[a], Ying Wu[a], Shang Wu[a], MingjiaZhu[a], Jianquan Liu[a,2], Eviatar Nevo[e,2], and Pilar Catalan[c,2]

[a] State Key Laboratory of Grassland Agro-Ecosystem, College of Ecology, Lanzhou University, Lanzhou 730000, China; [b] State Key Laboratory for Animal Disease Control and Prevention, College of Veterinary Medicine, Lanzhou University, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences,Lanzhou 730000, China; [c] Escuela Politecnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain; [d] University of Tel-Aviv, Tel-Aviv 6997801, Israel; [e] Institute of Evolution, University of Haifa, Mount Carmel, Haifa 3498838, Israel.

Corresponding author: Jianquan Liu, EviatarNevo, Pilar Catalán

Email: liujq@nwipb.cas.cn; nevo@research.haifa.ac.il; pcatalan@unizar.es;.

This PDF file includes:

Supporting Information Text

Materials and methods

Plant material

B. stacei seeds were collected from "Evolution Canyon I (ECI) " , located at Lower Nahal Oren, Mount Carmel, Israel (32°43′N; 34°*58′E*) (1). Each individual' seeds were separately collected, and a random selection of individuals was made from 'African slope (AS) and European slope (ES)'. Individuals from both slopes were selected in different plots through spatial transects covering all previously described microsites (AS: 1-3; ES: 5-7) at ECI for other organisms (DataSet S10) except site 4. In total, seeds of 25 and 16 individuals from AS and ES were respectively collected. We further collected seeds of five individuals out of ECI but within Israel (DataSet S1). The seeds of each individual were separately marked, germinated and grown in a growth chamber (providing 14 h light/10 h dark cycles and a constant 23± 2 °C) using Pindstrup Substrate (Pindstrup Mosebrug, Pindstrup, Denmark, www.pindstrup.dk). The fresh leaves of one seedling from each individual were separately collected for genome-resequencing. The seedlings of three randomly selected different individuals from each slope were used for physiological, transcriptomic and metabolic analyses.

Genome sequencing and assembly for Bsta-ECI

Representative *B. stacei* (Bsta-ECI) genome was generated and assembled from the seedling leaves of one individual from AS at ECI using a PacBio Sequel2 platform. High-quality genomic DNA was extracted from fresh leaves using a DNeasy Plant Mini Kit (Qiagen, USA) to construct 15-kb SMRTbellTM libraries prior to circular consensus sequencing (CCS). 25.01 Gb of CCS reads were generated for the Bsta-ECI genome. Illumina paired-end libraries were constructed from the same plant for long-read and sequenced by an Illumina HiSeq X Ten machine, totaling 17.24 Gb of reads' yield. Hi-C libraries were constructed from chromatin, fixed with formaldehyde in the nucleus, and cross-linked DNA was digested with DpnII. Approximately 151.75 Gb of Hi-C raw reads were obtained using an Illumina HiSeq X instrument (Table S1). Genome size was estimated using K-mer methods based on Illumina short reads. K-mer depth distributions were generated with KMC3 (2) and genome sizes were calculated with genomescope v2.0 with default parameters (v1.94) (3).The HiFi reads were assembled using NextDenovo (v2.4, https://github.com/Nextomics/NextDenovo) with default parameters except for 'read_type,' which was set to 'hifi'. Hi-C technology was applied to anchor contigs to chromosomes of the Bsta-ECI genome. Low-quality Hi-C reads were filtered using Fastp (v0.20.0) (4) with '-q 20 -5 -3 -w 10' settings, and high-quality filtered reads were mapped to the corresponding draft genomes using Juicer (v1.9.9) (5). After filtering and removing misaligned reads, the genome sequences were clustered and oriented using 3D-DNA (v201008) (6) to generate the final chromosome-level assembly of Bsta-ECI. We employed Merqury (v1.3) (7) to evaluate genome assembly quality based on K-mer methods. In addition, all filtered short Illumina reads of ABR114 and Bsta-ECI were mapped to their corresponding genome assembly using BWA-MEM2 (v2.2.1) (8) to generate align bam files. SAMtools (v1.1) (9) and Bamdst (https://github.com/shiquan/bamdst) were used to count read's map ratios and coverage for each genome assembly. Contig NG (X) (10) values were calculated as the reference and plotted using custom Perl scripts. Genome

completeness was assessed using BUSCO (v3.0.2) (11) with the embryophya_odb10 dataset, and LTR Assembly Index (LAI) values were generated with LTR-retrieve (v2.9.0) (12) with default parameter.

Genome annotation

Repetitive DNA sequences and gene models were identified and predicted following a previous study (13). Tandem repeats were annotated using TRF (v4.09b) (14) and transposable elements were identified using RepeatModeler (v2.0) (15), LTR-retrieve (v2.8.5) (12), RepeatMasker (v4.0.7) (16), and RepeatProteinMasker (included in the RepeatMasker package) based on the published Repbase (v20181026) (17). We merged all results and removed redundancies using the subcommand 'merge' in Bedtools (v2.29.1) (18) for the final repeat annotation. The strategy for gene prediction combined *ab initio*, homology-based, and transcriptomic methods. Augustus (v3.2.3) (19) was used for *ab initio* gene model predictions with a custom training dataset created by BUSCO (v3.0.2) (11). For homology-based prediction, we collected genome data of five representative plant species: *Arabidopsis thaliana* (GCF_000001735.4), *Oryza sativa* (GCF_001433935.1), *Sorghum bicolor* (GCA_000003195.3), *Zea mays* (GCA_000005005.6), *Brachypodium distachyon* (JGI. v3.1), and the published reference genomes of *B. stacei*-ABR114 and *B. hybridum*-ABR113 (20), and aligned their protein to the newly assembled Bsta-ECI genome to search for orthologous gene structures, which was integrated by GeMoMa (v1.7.3) pipeline (21) . Besides, alignment- and de novo-based methods were applied to transcriptome-based gene predictions. For the alignment-based method, we employed HISAT2 (v2.2.1) (22) to map RNA-Seq data to the sequenced genome, then Stringtie (v2.1.6) (23) and Transdecoder (https://github.com/Trans-Decoder/TransDecoder) were used to generate the annotation file. For the *de novo*-based method, we used Trinity (v 2.9.1) (24) and the PASA-pipeline (v2.4.1) (25) to assemble transcripts and obtain gene feature files (GFF). All gene models predicted by these approaches were integrated through EvidenceModeler (EVM; v1.1.1) (26). We used BUSCO-3.0.2 to assess the completeness of gene model predictions using the embryophya_odb10 dataset. For functional annotation of gene models, we employed InterproScan (v5.36) (27) to annotate the predicted gene models completely. The Blast program was used to search for homeologous candidates in three public databases: Swiss-Prot (28), NR (29), and TrEMBL. Gene Ontology (GO) term annotations combined results from InterProScan, the EggONG (30) database, and Blast2GO (v6.0) (31) pipeline. The iTAK program (web version) (32) was used to identify transcript factors. The Kyoto Encyclopedia of Genes and Genomes (KEGG), EggONG was used for KEGG pathway analyses.

Whole-genome resequencing and variant calling

A total of 46 *B. stacei* individuals were used for genome-resequencing. They comprise 25 and 16 individuals for AS and ES at ECI and five ones outside of ECI in Israel. The DNA extractions and experiments followed the previous study (ref. 22 in the main text). Illumina paired-end libraries were constructed and sequenced using an Illumina HiSeq X Ten machine. Three additional *B. stacei* genomes (SRR11671585, SRR1802178, ERR4835442) and one *B. distachyon* genome (SRR191989) were downloaded from the NCBI database and used as outgroups.

Low-quality reads were filtered and discarded using Fastp (v0.20.0) (4). Multiple alignment files were generated with BWA-MEM2 (v2.2.1) (8) using the Bsta-ECI genome assembly as reference and then sorted with SAMtools (v1.1) (9). PICARD (v.2.23.4, http://broadinstitute.github.io/picard/) was applied to remove PCR duplications. The quality of alignments near indels was improved using the 'IndelRealigner' and 'RealignerTargetCreator' modules of the Genome Analysis Toolkit (GATK) package (v3.8.1) (34). We used the 'HaplotypeCaller' module of GATK for single nucleotide polymorphism (SNP) calling to generate a gvcf file for each sample, and the 'GenotypeGVCF' module of GATK for joint genotyping across samples. Then, we applied multiple filtering steps to reduce false positive rates and obtain high-quality SNP data for downstream analysis. First, GATK the 'VariantFiltration' module was applied to filter multi-sample SNPs with the following parameters: 'QD < 2.0 || FS > 200.0 || ReadPosRankSum < −20.0', and 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0' for indel and SNP filtering, respectively. Then, we used a custom Perl script and VCFtools (v0.1.17) to remove low-quality SNPs including non-biallelic SNPs; SNPs with low (< 1/3 of chromosome average) or high (> 3-fold chromosome average) depth, or < 10 quality (GQ) scores for genotype calls; SNPs missing in > 20% of samples; SNPs within 5 bp of indels or in repetitive regions; and SNPs with minor allele frequencies of < 0.01. Filtered SNP data were annotated with SnpEFF (35). We employed Manta (v1.60) (36) with default parameters for analysis of genome structural variants (SVs), detecting SVs using realigned BAM files, and filtering low-quality SVs with SURVIVOR (37). We applied TEPID (v 0.10) (38) with default pipeline parameters for transposable element polymorphism (TEP) analysis.

Population structure analysis

The full nuclear genotypic SNP data set (722,351 SNPs) was transformed into a multi-aligned FASTA format using a custom Perl script, and used for phylogenetic, population structure and demography analyses. Furthermore, in order to obtain a data matrix consisting only of neutral markers, a reduced genotypic SNP data set (25,471 SNPs) was generated from the full data set using four-fold degenerate synonymous sites extracted using FxTools (https://github.com/BGI-shenzhen/FxTools), which was also used to perform phylogenetic and population structure analyses. A maximum likelihood phylogenetic tree was constructed using IQTree (v 1.6.12) (33) imposing the best-fit nucleotide substitution model according to ModelFinder sub-parameter of IQTree. Branch support for the best tree was estimated through 1,000 ultrafast bootstrap replicates (command: "iqtree -s snp.fa -pre snp -nt 40 -bb 1000 -m MFP"). Population structure of *B. stacei* AS and ES populations was analyzed using ADMIXTURE (v1.3.0) (34) with K values set from 2 to 4, in each case with 100 bootstrap replicates. PCA was performed with the 'smartpca' packages in EIGENSOFT (v7.2.1) (35) using default parameters. Pairwise relatedness between samples was calculated using KING (v1.4) (36). The Neighbor-Joining tree of SVs was constructed with the Phylip package (37). For TEPs, we encoded the presence of a TE polymorphism as 1 in one individual and its absence as 0, thereby generating a M × N matrix, where M represents the number of detected TE polymorphic elements and N represents the number of samples. Then, we performed Principal Component Analysis (PCA) analysis based on this 0-1 coding matrix using the 'prcomp' function of the R 'stats' package.

Genetic parameters estimation

The nucleotide genetic diversities (θ_π) of the *B. stacei* AS and ES populations were estimated with VCFtools (v0.1.17)(38) using the non-overlap 10k bp window for each population. Whole-genome genetic diversity was defined as the mean value of $\theta\pi$. Linkage disequilibrium (LD) was calculated and plotted based on the coefficient of determination (r^2) using PopLDdeacy (v3.41) (39). Selfing rates were estimated using methods described previously (40). Briefly, we employed VCFtools(38) to transform vcf files into binary files which were analyzed with PLINK (v1.9) (41) to calculate inbreeding coefficients (F_{IS}), and selfing rates (s) were obtained from the formula $s = 2F_{IS}/(1 + F_{IS})$ (40). The recombination rates were calculated by FastERPP (http://www.picb.ac.cn/evolgen/softwares/FastEPRR.html) using the non-overlap 10k─bp window.

Demographic history inference

Site frequency spectra (SFS) can be distorted when even entries are greater than odd entries, due to excess homozygosity in inbred populations. High levels of inbreeding generate frequency spectra with a zig-zagging pattern between adjacent entries. According to a previous study (40) and our estimates, *B. stacei* has a highly selfing mating system, therefore we inferred the demographic history of the *B. stace*i AS and ES populations using the *∂a∂i* (42) inbreeding module. We first masked the genic and repeat regions of the SNP data to exclude the influence of natural selection and mapping errors. Then we fitted parameters for seven hypothetical demographic models ('no_mig', 'sym_mig', 'anc_asym_mig', 'asym_mig', 'sym_mig_twoepoch', 'sec_contact_asym_mig', and 'asym_mig_twoepoch') using the dadi_pipeline (v3.1.5) (43) and performed consecutive rounds of optimization for all models. For each model, we ran 100 rounds with default settings in the dadi_pipeline and recorded parameter estimates from the best scoring round. We compared fits of the models to the data using the Akaike information criterion (AIC) and residuals plot

.

Genetic divergence and selection

The Fixation statistic (F_{ST}) and mean pairwise nucleotide differentiation (D_{XY}) of *B. stacei* AS vs ES populations were estimated using pixy (44) with a 10 kb window size. The high divergent genomic islands were identified by a permutation-empirical combined method (45,46) This method avoided statistical power biases caused by fixed thresholds for non-normal F_{ST} distribution. Briefly, we first excluded windows containing fewer than 10 SNPs and considering windows within the top 5% F_{ST} values as outlier windows. Then, for each window, an equivalent number of SNPs was randomly rearranged from the entire genome-wide SNPs, and the corresponding F_{ST} values were then calculated. This permutation process was repeated 5,000,000 times in total generating null distributions of F_{ST} values. Subsequently, the p-values of the assumed outlier windows were determined by comparing them with the null F_{ST} distributions. All p-values underwent correction using the False Discovery Rate (FDR) method. Outlier windows were defined as those belonging to the experimental set with FDR values lower than 0.01. The overlaps of outlier F_{ST} and D_{XY} windows were considered divergent regions of reduced gene flow and/or fixed ancient polymorphism. For selection analysis, we applied the Hudson-Kreitman-Aguadé (HKA) test (47) to detect genes saturated by recent selective sweep, indicative of natural selection pressure. HKA tests for each gene were

performed as previously described (45). Briefly, we scanned the coding region of all genes to count numbers of polymorphic sites and fixed sites ($F_{ST} > 0.95$) of each gene ($Gene_{poly}$ and $Gene_{fix}$, respectively) and of all genes ($Genome_{poly}$ and $Genome_{fix}$) between the AS and ES populations. Then, HKA tests were applied by comparing $Gene_{poly}/Gene_{fix}$ ratios to the genome-wide average. The null hypothesis $Gene_{poly}/Gene_{fix} = Genome_{poly}/Genome_{fix}$ was tested using the Pearson's chi-square test with a 2x2 contingency table (48).

The plastome phylogenies

The *B. stacei* chloroplast genomes (plastomes) of the ECI samples and several other accessions (Cef2, TE4.3, other samples from Israel and Bd30) were assembled using Getorganelle (v1.7.1)(49). All plastomes were annotated using PGA (v3.0) (50) and three reference chloroplast genomes (NC_011032, NC_036836, NC_036837). Coding DNA sequences (CDS) of plastome genes and whole plastome sequences were aligned using MAFFT (v6.864b) (51), then the maximum-likelihood (ML) trees of concatenated plastome genes and the whole plastome were constructed with IQ-TREE (v 1.6.12) (33) with the '-bb 1000 -m MFP' parameters. Plastome haplotype frequencies were calculated with DNAsp (v6.0) (52) and a statistical parsimony haplotype network was constructed with Popart (v1.7) (53). Only representative samples in each leaf cluster of Fig 2A were selected to show in tree and haplotype network analysis.

Drought stress, physiological, transcriptomic and metabolomic analysis

Seeds from three AS and three ES individuals were respectively germinated and grown for 14 h light/10 h dark cycles and a constant 23± 2 °C using Pindstrup Substrate (Pindstrup Mosebrug, Pindstrup, Denmark, www.pindstrup.dk). Three randomly selected seedlings for each individual were used for all analyses. The drought treatment experiment started in the third week after germination (21d), and was monitored through soil moisture measurements (field water capacity). We assigned seedlings of all individuals to two treatment groups: a drought treatment in which water was withheld for one week, and a well-watered group that was watered every two days for the duration of the experiment. Leaf and root samples from well-watered and drought-treated plants were collected for physiological parameter measurements and for transcriptomic and metabolomic analysis. Five physiological parameters were measured with a LI-6800 infrared gas analyzer (Li-Cor Inc., Lincoln, NE, USA): assimilation rate (A), stomatal conductance to water vapor (gsw), intercellular CO2 concentration (Ci), and transpiration rate (E). Water use efficiency (WUE), was calculated using the formula WUE = A/E*1000. For root growth assay, seeds randomly selected from individuals of the two populations were subjected to a 10-minute treatment with 10% sodium hypochlorite, followed by a 7-day incubation at 4°C. After 1 day of germination, the seeds were transferred to MS culture medium containing 0, 5, and 40 μM ABA, respectively. Root lengths were measured after 5 days' growth.

Total RNA was extracted from the mixed leaf or root tissues of three seedlings from each individual using the QIAGEN RNeasy plant mini kit, and sequenced using a DNBseq-T7 system following RNA library construction with oligo dT methods. A total of 349.608 Gb RNA-seq paired-end Illumina reads were obtained across all samples. Transcriptome analysis was conducted using the 'HISAT2-Stringtie-DESeq' pipeline (54). RNA-Seq reads were mapped to the representative genome Bsta-ECI with HISAT2 (v2.2.1) (22) using the '--dta -p 20' parameter.

Then we used Stringtie (v2.1.6) (23) to convert bam files to transcript gtf files and calculate gene expression abundances (FPKM and TPM), using our gene feature files as guides. Stringtie outputs were converted to a read count matrix using the Python script 'prepDE.py3'. Differential expression patterns were analyzed with the R package 'DESeq2' (55), and using 'Log2foldchange (absolute value) > 1' and 'P.adjust (BH correction methods) < 0.01' as thresholds for differentially expressed genes (DEGs) in each comparison.

A portion (20 mg) of each root and leaf sample was ground to a powder from which metabolites were extracted using 70% aqueous methanol. Samples were then centrifuged at 4°C and 13,000 rpm for 10 min and supernatants were collected for Liquid Chromatography – Mass Spectrometry (LC-MS) analysis. LC-MS analysis was performed using an ACQUITYUPLCI-Class system coupled with a VION IMS QTOF mass spectrometer (both from Waters Corporation, Milford, USA), operated in both electrospray ionization (ESI) positive and negative modes. Raw LC-MS data were processed with Progenesis QI V2.3 software (Nonlinear, Dynamics, Newcastle, UK). Metabolites were identified using the precise mass-to-charge ratios (M/z), secondary fragments, and isotopic distributions using multiple databases for qualitative analysis. The retrieved data were then further processed by removing any peaks with a missing value (ion intensity = 0) in more than 50% in groups, by replacing zero values by half of the minimum value, and by screening the data according to the qualitative results of the compound. Compounds with resulting scores below 36 (out of 60) points were also deemed to be inaccurate and removed. The positive and negative ion data were combined for downstream analysis. Orthogonal Partial Least-Squares-Discriminant Analysis (OPLS-DA) and Partial Least-Squares-Discriminant Analysis (PLS-DA) were employed to identify metabolites that differed between groups. Variable Importance of Projection (VIP) values obtained from the OPLS-DA model were used to rank the overall contribution of each variable to group discrimination. A two-tailed Student's T-test was used to assess the significance of differences in metabolite composition between groups. Differential metabolite compositions were defined as those with VIP values greater than 1.0 and p-values less than 0.05.

Gene ontology (GO) and KEGG pathway enrichment analysis.

GO annotation for the *B. stacei* genes was assigned as previously described and GO enrichment analysis was performed with GOATools (v1.1.6) (56) with BH correction methods. KEGG pathway enrichment was conducted with ClusterProfile (v3.18.1) (57). The results of GO enrichment and KEGG pathway enrichment were visualized with ReviGO (http://revigo.irb.hr/) and a custom R script.

Statistical analysis.

We used R (v4.0; https://www.r-project.org/) for several statistical analyses, and the R packages ggplot2 (v3.3.5) packages (https://ggplot2.tidyverse.org/) was used to visualize the results.

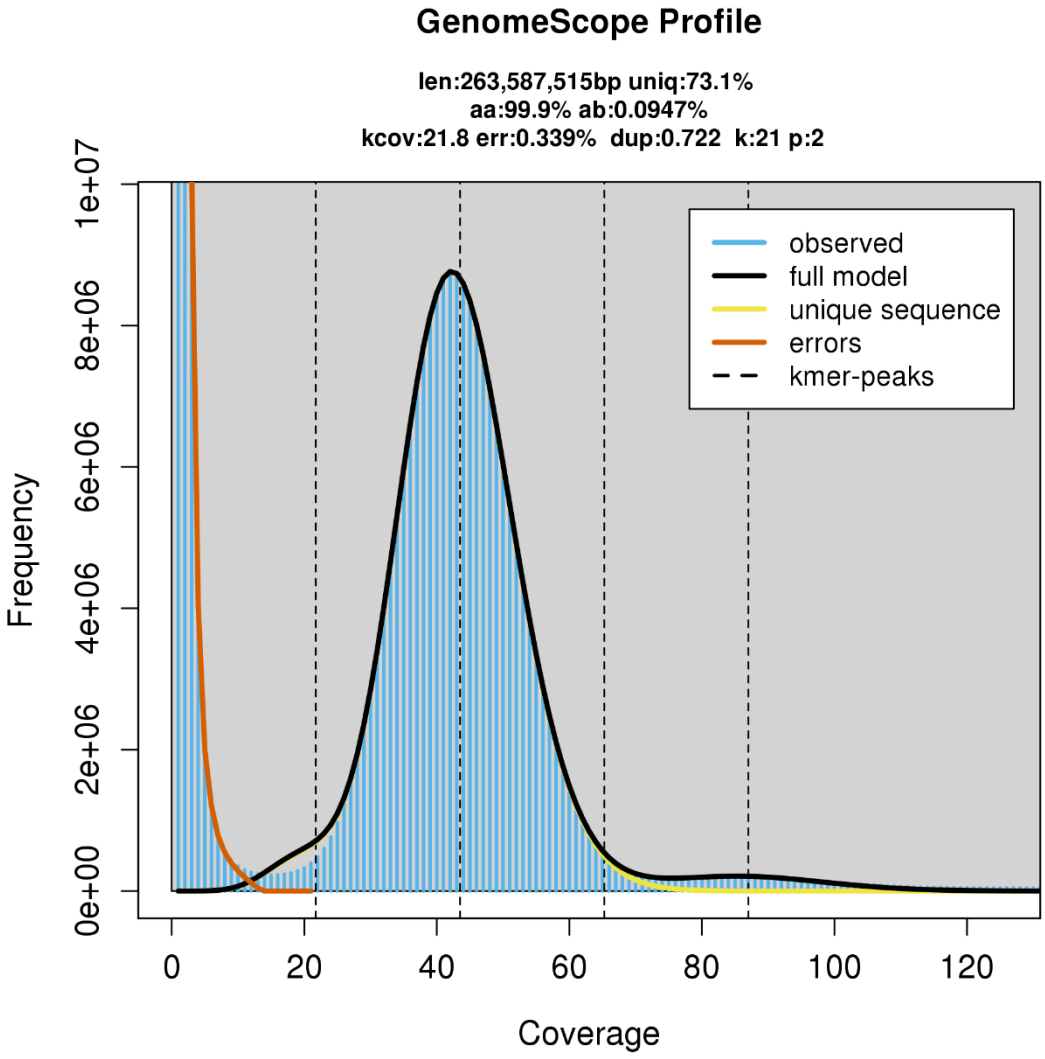

Figure S1. Genome survey of the *B. stacei* Bsta-ECI accession from Evolution Canyon I (see Table S1 for additional information)**.** K-mer size was set at 21 and the default parameters were set with GenomeScope2.

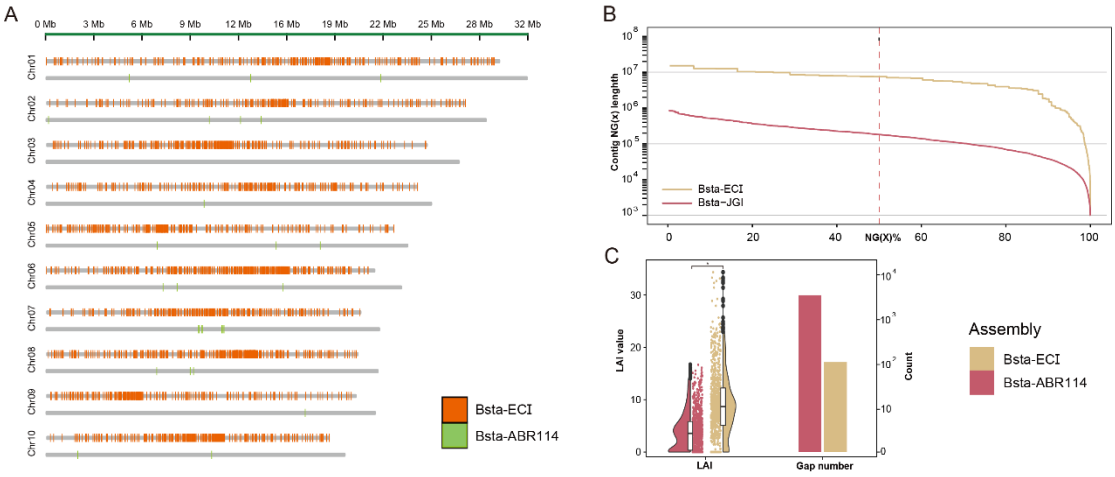

Figure S2. Compared genomic features of the new Bsta-ECI assembly from Evolution Canyon I and *B. stacei* reference genome ABR114 obtained from Illumina sequencing at the Joint Genome Institute (Phytozome: https://phytozome-next.jgi.doe.gov/). **A)** Chromosomes length statistics of genome assembly, gap positions with respect to genomes were marked as color bars. **B)** Overview of assembly contig lengths comparing NG values, from 1 to 100%, and the contig length (in bp) for particular thresholds are shown on the y-axis. Calculated NG contig lengths range on a log scale. The dashed vertical line indicates the NG50 contig length. **C)** Comparison of gap numbers and the LTR Assembly Index (LAI) among assembly versions, *p-value < 0.01 (t-test).

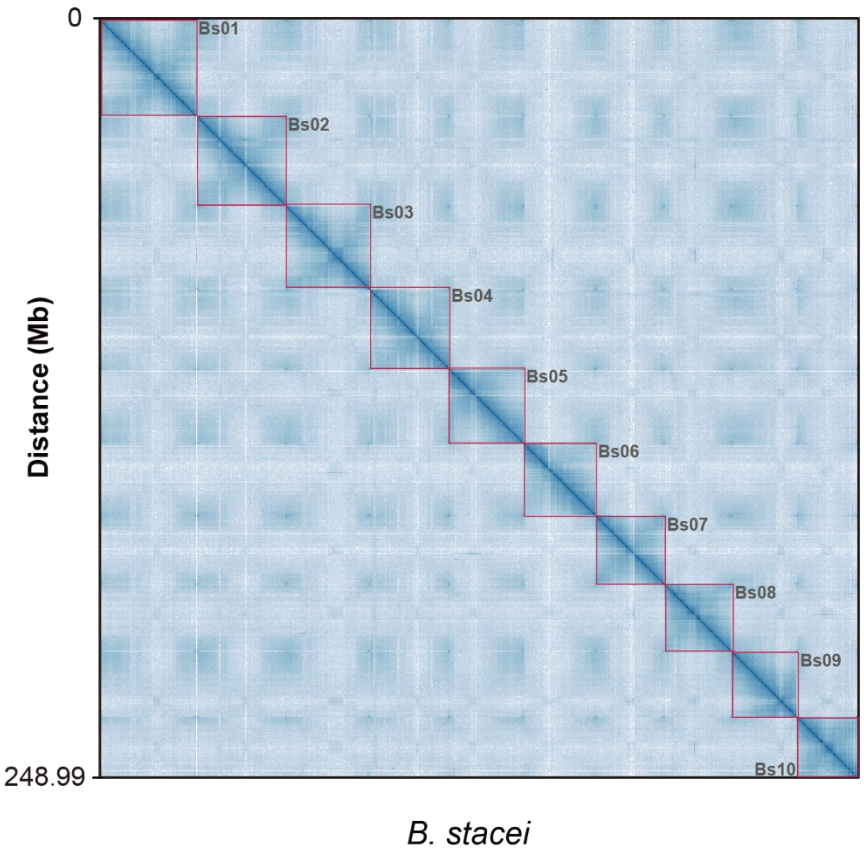

Figure S3. Newly assembled Bsta-ECI genome. Hi-C interaction densities heatmap between contigs; its 10 chromosomes (Bs01-Bs10) are separated by red boxes.

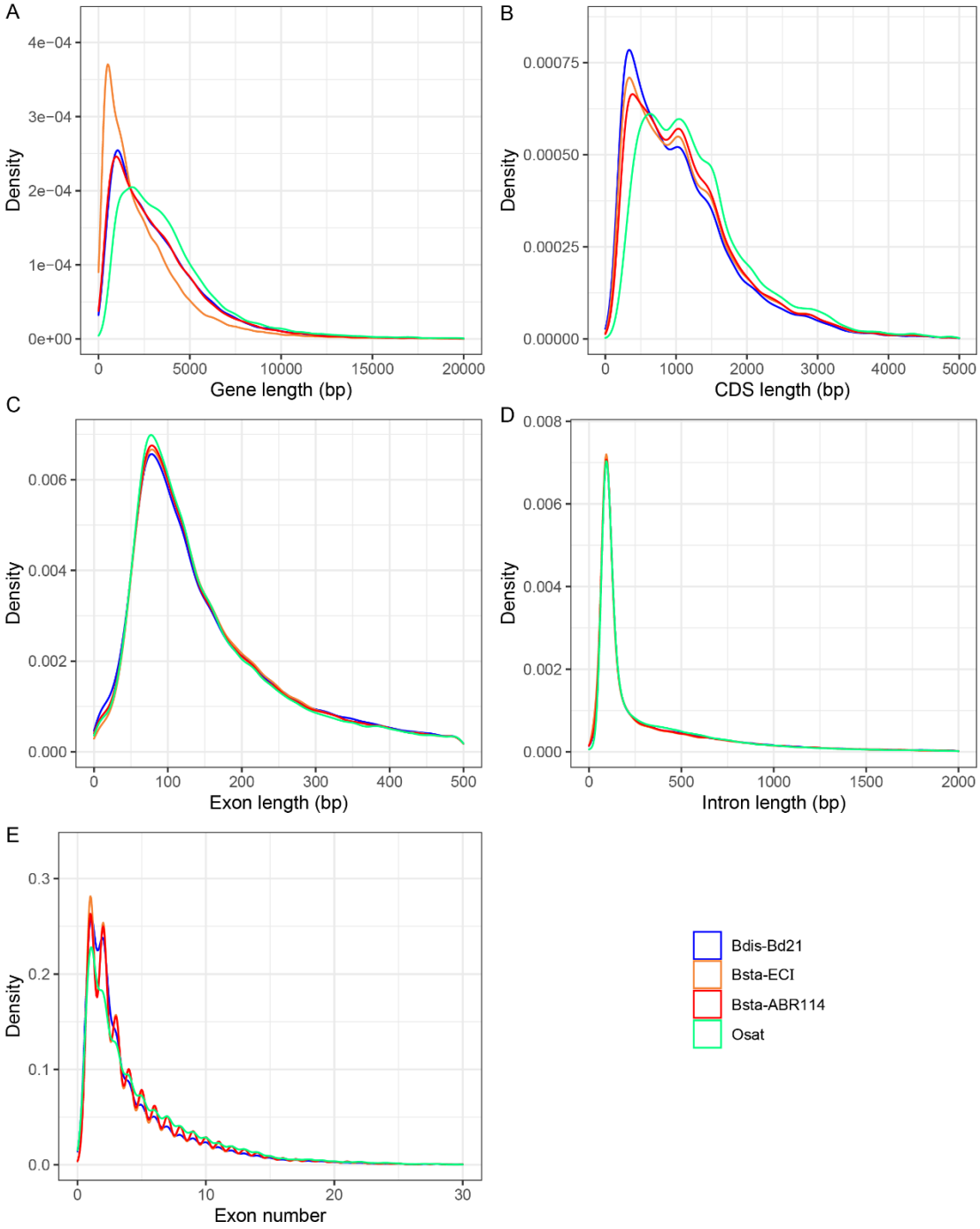

Figure S4. Comparison of gene model structure characteristics in the newly assembled Bsta-ECI genome to those in other assemblies (reference genomes of *B. stacei* ABR114 v. 1.1, *B. distachyon* Bd21, generated by the Joint Genome Institute (JGI), see Gordon et al. 2020) and *Oryza sativa* (Osat). **A)** gene length; **B)** CDS length; **C)** exon length; **D)** intron length; **E)** exon number.

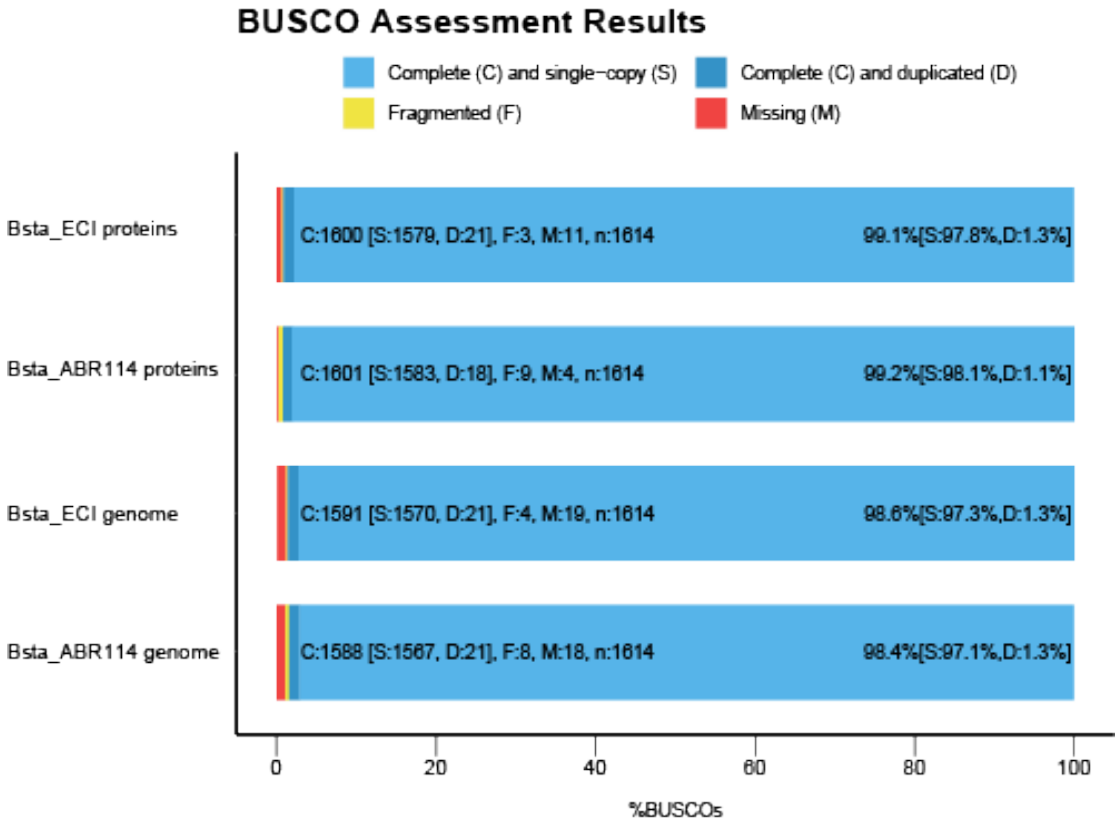

Figure S5. BUSCO gene assessments of genome assemblies. Color codes for gene types are indicated in the chart. Bsta-ECI genome, newly assembled *Brachypodium stacei* genome from Evolution Canyon I; Bsta-ABR114, *B. stacei* reference genome (Phytozome; https://phytozome-next.jgi.doe.gov/). BUSCO runs in "genome" and "proteins" modes.

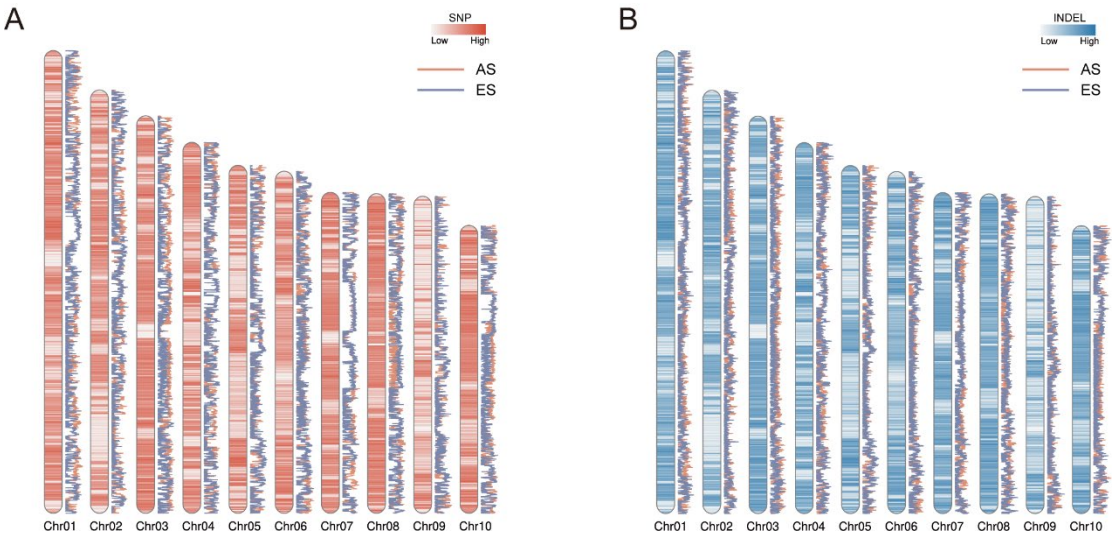

Figure S6 The distribution of SNPs (red) and INDELs (blue) obtained after mapping the genome data of the 41 resequenced *B. stacei* ECI individuals to the reference Bsta-ECI genome across the 10 chromosomes of *B. stacei* (using 10kb non-overlapped window) and for each population separately. **A**) The SNP density across the 10 chromosomes, the density of total SNPs (chromosome heatmap), SNPs only present in AS (salmon line) and in ES (purple line). **B**) The INDEL density across the 10 chromosomes, the density of total INDEL (chromosome heatmap), INDELs only present in AS (salmon line) and in ES (purple line). The densities of SNPs and INDELs are similar for the two populations, indicating similar coverages for the called genotypes.

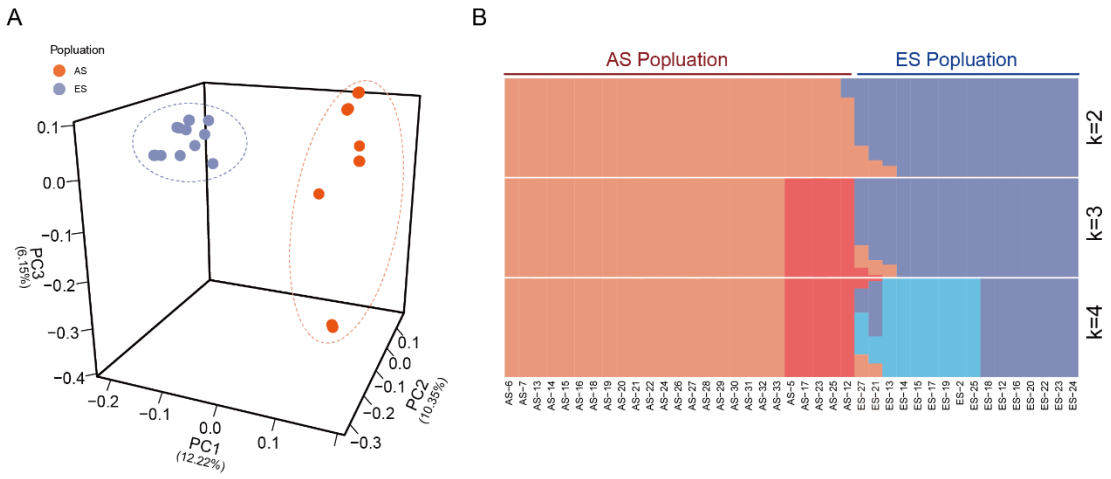

Figure S7. Population structure analysis of 41 *B. stacei* resequenced individuals from Evolution Canyon I (ECI) based on nuclear genome SNP data**. A)** Tridimensional plot of Principal Component Analysis (PCA). The PCA1 axis (accumulating 12.22% of the variance) clearly separated the two populations from the African Slope (AS) and the European Slope (ES). **B)** Population structure bar plot retrieved for K=2, 3, and 4 hypothetical populations. The optimal structuring corresponded to K=2, which also separated the AS and ES populations; another less-optimal Ks showed some substructuring within AS (K=3) and AS and ES (K=4).

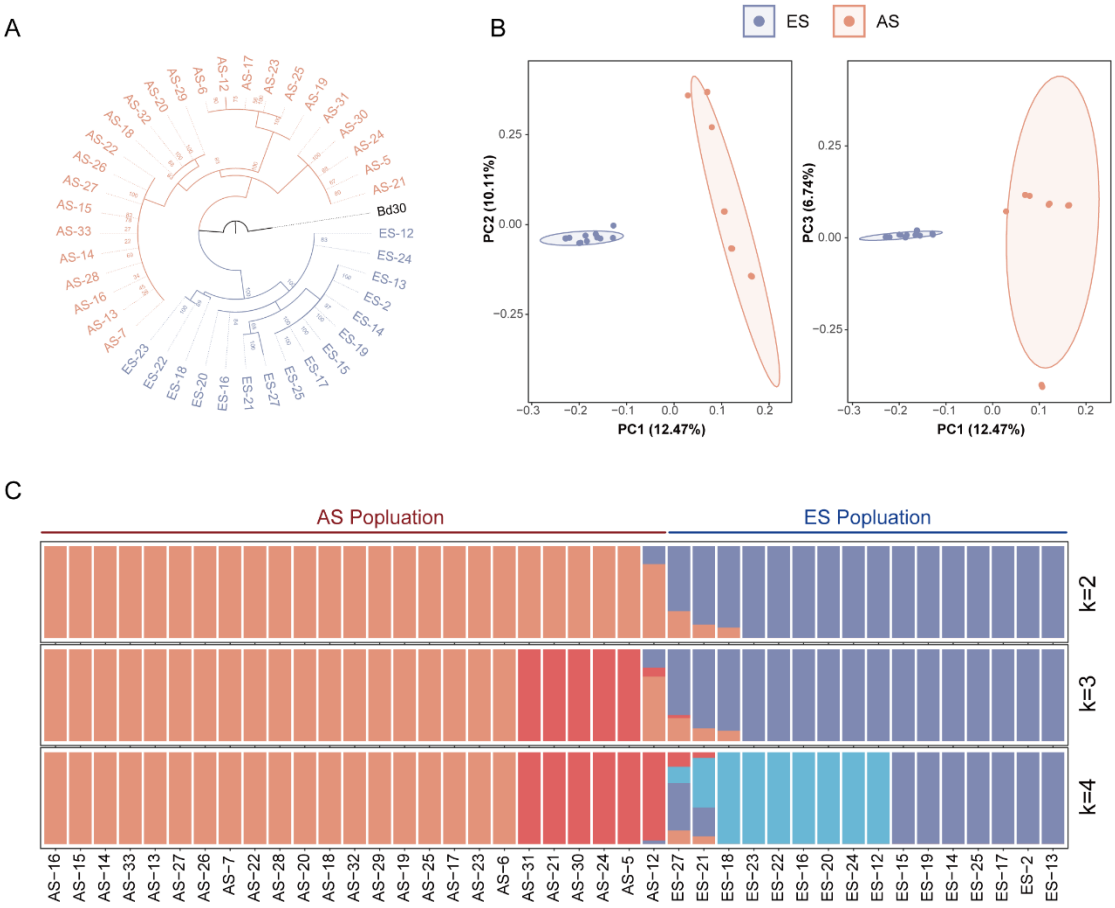

Figure S8. Population structure analysis of 41 *B. stacei* resequenced individuals from Evolution Canyon I (ECI) based on four-fold degenerated SNP sites. **A)** Maximum-likelihood phylogenetic tree; **B)** Bidimensional PCA plots of PC1-PC2, and PC1-PC3 axes; **C)** Population structure bar plot retrieved for optimal K=2, and less-optimal K=3, and K=4 hypothetical populations. The phylogenetic and genetic structure analyses based on four-fold degenerated SNP sites rendered similar results to those obtained with the full SNP data set (see Fig. S7).

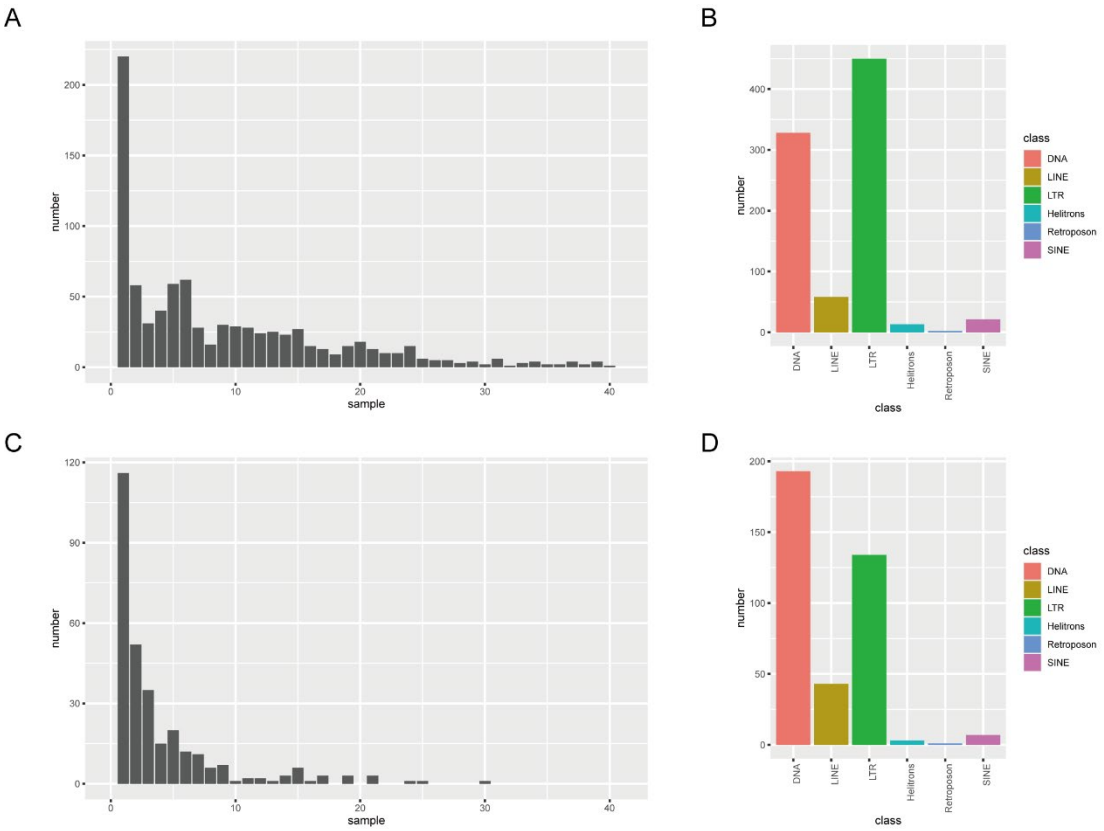

Figure S9. Distribution and classification of transposable elements' polymorphisms (TEPs) found in the 41 *B. stacei* genomes studied from ECI. **A)** Distribution of deletions among all 41 genomes relative to the Bsta-ECI genome. The x-axis indicates the number of samples showing deletions and the y-axis the number of polymorphic transposable elements. **B)** Classification of polymorphic repeat elements showing deletions with respect to the Bsta-ECI genome. **C)** Distribution of insertions among all 41 genomes relative to the Bsta-ECI genome; x and y axis indicate the number of samples showing insertions and the number of polymorphic transposable elements, respectively. **D)** Classification of polymorphic repeat elements showing insertions with respect to the Bsta-ECI genome.

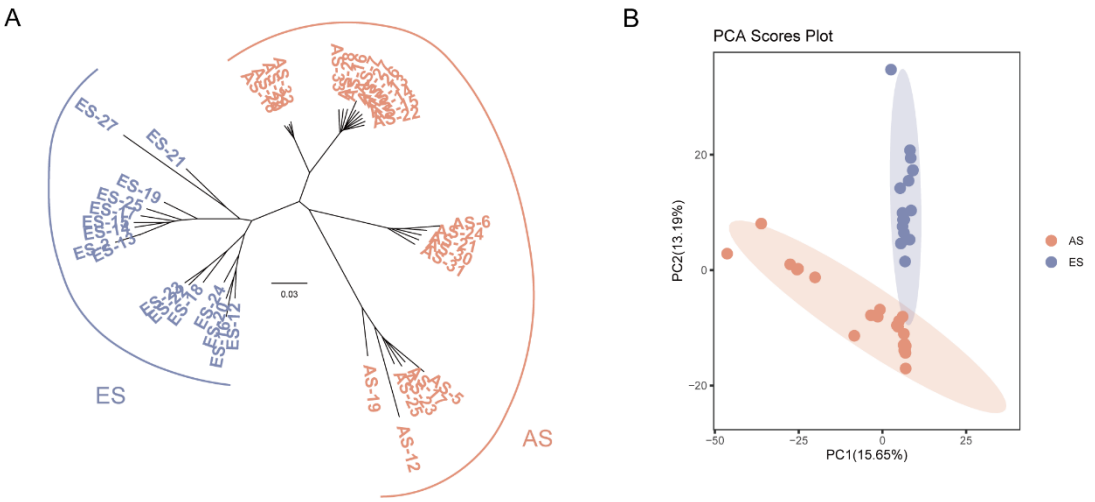

Figure S10. Population structure analysis of 41 *B. stacei* resequenced individuals in ECI using Structure Variants (SVs) and Transposable Elements Polymorphisms (TEPs) data obtained from the 41 resequenced genomes. **A)** Neighbor joining tree based on SVs data (see Supplementary Table S12) with variants coded as binary characters; **B)** Bidimensional PCA plot based on TEPs coded by frequency classes as binary data. Both analyses separate the samples from the African Slope (AS) and the European Slope (ES) populations.

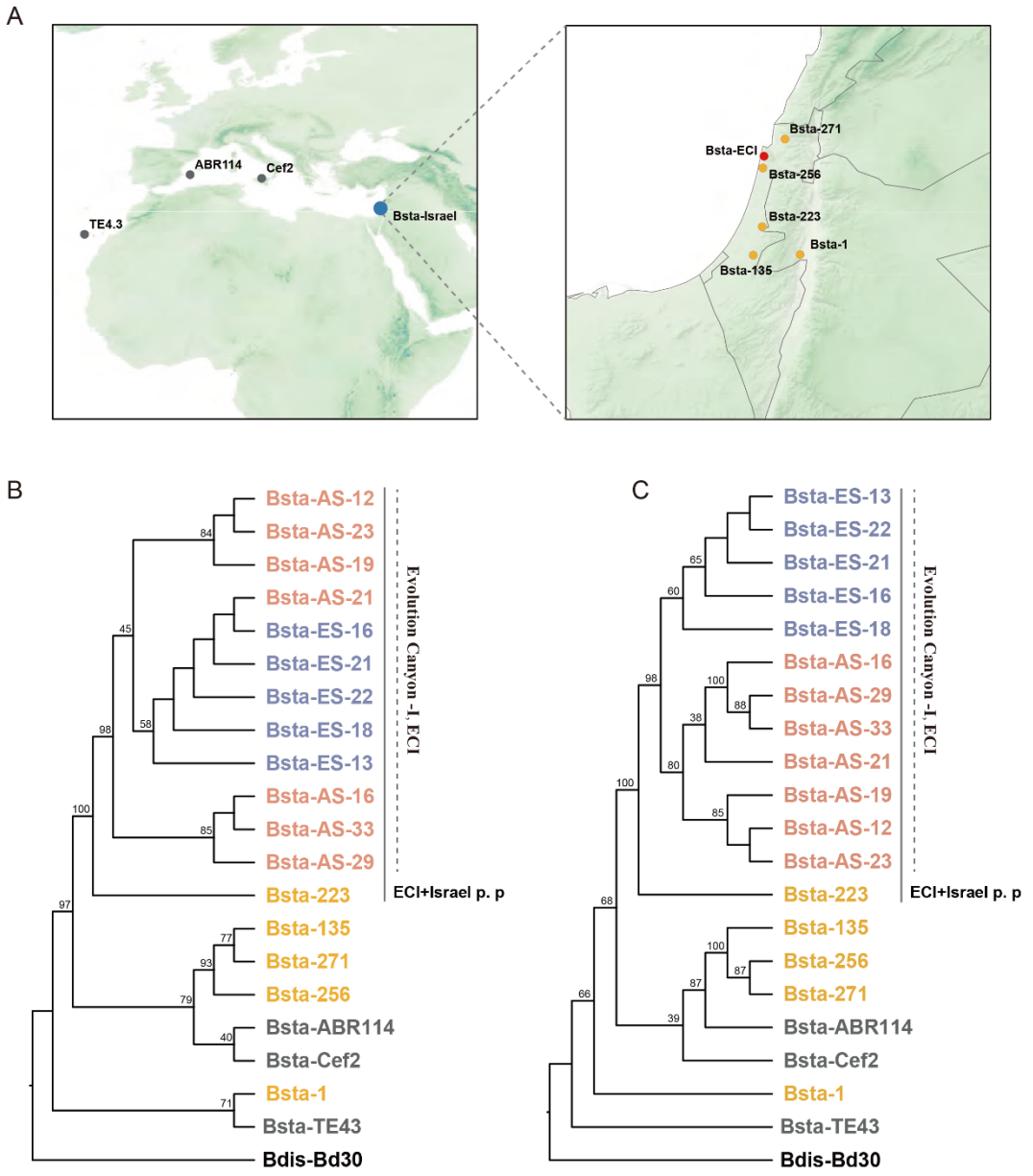

Figure S11. The Geographic distribution of *B. stacei* samples used in this study and maximum-likelihood phylogenetic trees based on whole plastome sequences; numbers at nodes indicate bootstrap support values. **A)** the location of samples; **B)** concatenated plastome genes tree**; C)** whole plastome tree**.** Clades: AS, African Slope; ES, European Slope; ECI, Evolution Canyon I; ECI+Israel p. p. (pro partim), ECI + Is-223. In the AS population, two polymorphic sites in 76 chloroplast genes were identified, and formed three haplotypes (GC[5 samples, 20%], GT[6 samples, 24%], and AC[14 samples, 56%]), while in the ES population, only one haplotype (GC) was observed (16 samples, 100%). Only representative samples in each leaf branch of Fig 2A were used to construct the plastome ML trees.

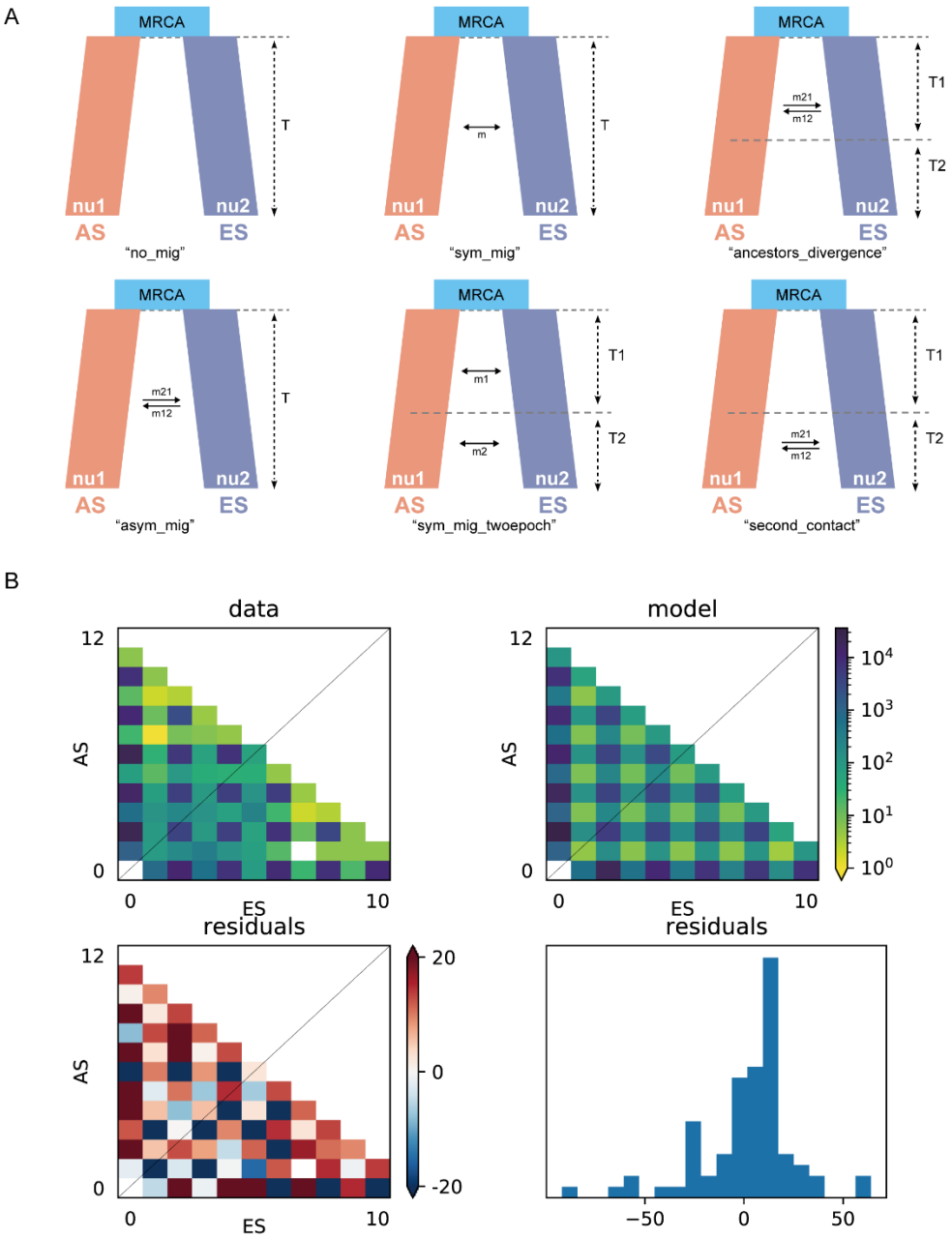

Figure S12. A) Six other alternative *∂a∂i* models used in the inference of the evolutionary demographic histories of the *B. stacei* AS and ES populations in ECI using forward simulation and residuals analysis. **B)** Demographic parameters (p0 = [0.3357; 0.1978; 3.1526; 1.7528; 1.542; 1.3721; 0.1131; 0.3788]) used to perform the residuals analysis of the best fit *∂a∂i* model (see Fig. 2C).

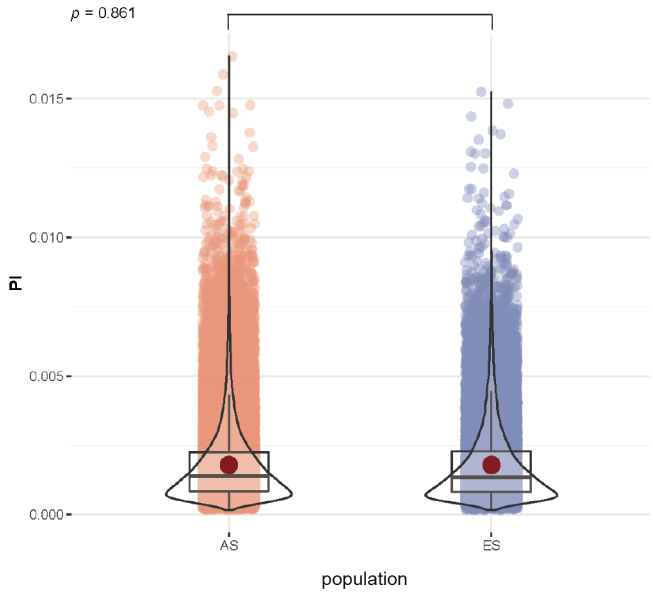

Figure S13. Population nucleotide diversity (θ_π) of the 41 studied *B. stacei* resequenced genomes at the ECI site. Distribution of $\theta\pi$ values for the AS and ES populations, calculated by 50kb window and 12.5kb step using vcftools. Both populations show similar mean θ_π values.

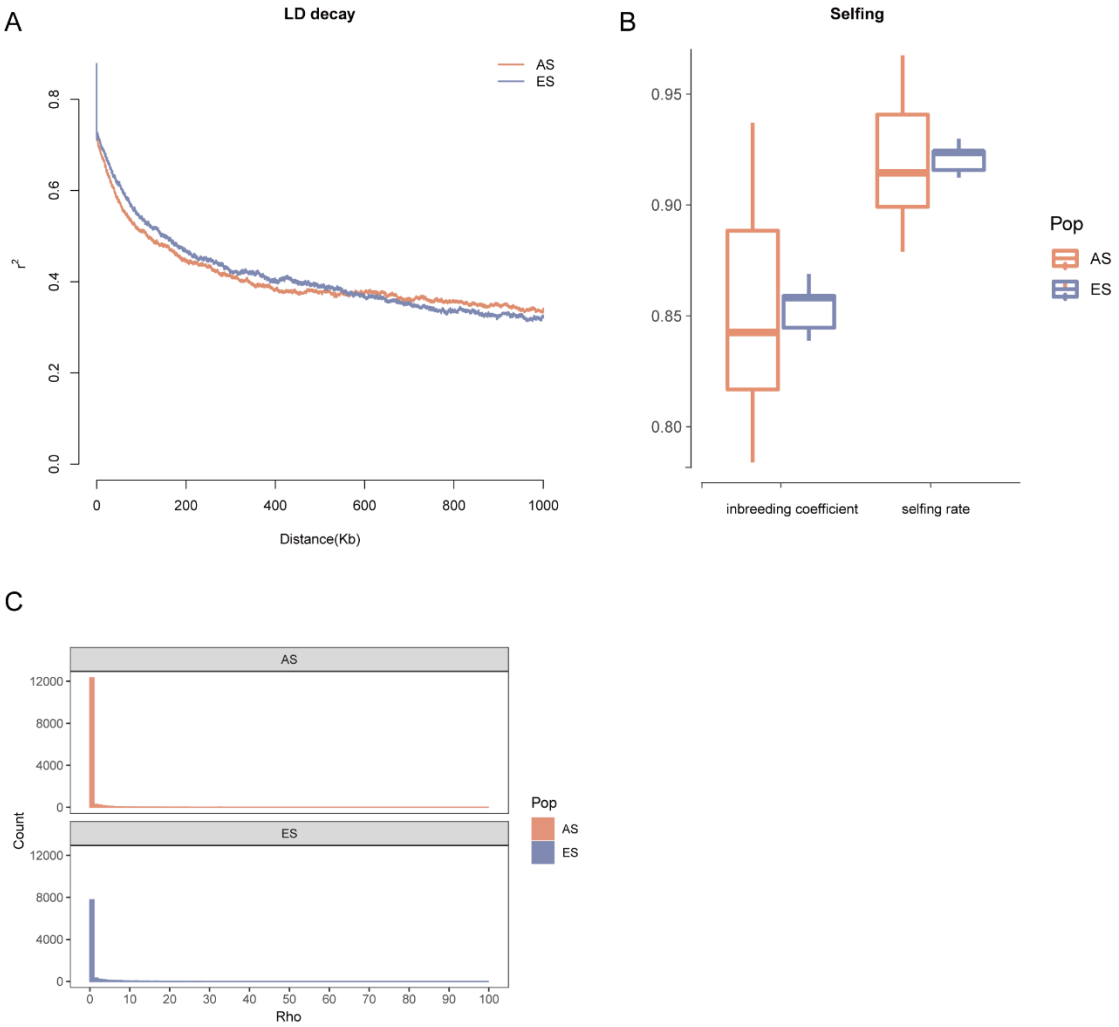

Figure S14. Linkage disequilibrium decay (LD) and inbreeding rates of 41 *B. stacei* resequenced genomes from the African Slope (AS) and European Slope (ES) populations at ECI. **A)** LD decay of the two AS and ES populations (x axis: physical distances (Kbp) of the individual genomes; y axis: r^2 value); **B)** box plots showing inbreeding coefficient values and selfing rate values for the AS and ES populations of ECI calculated with PLINK. **C)** The histogram plot shows a low recombination rate for the AS and ES populations.

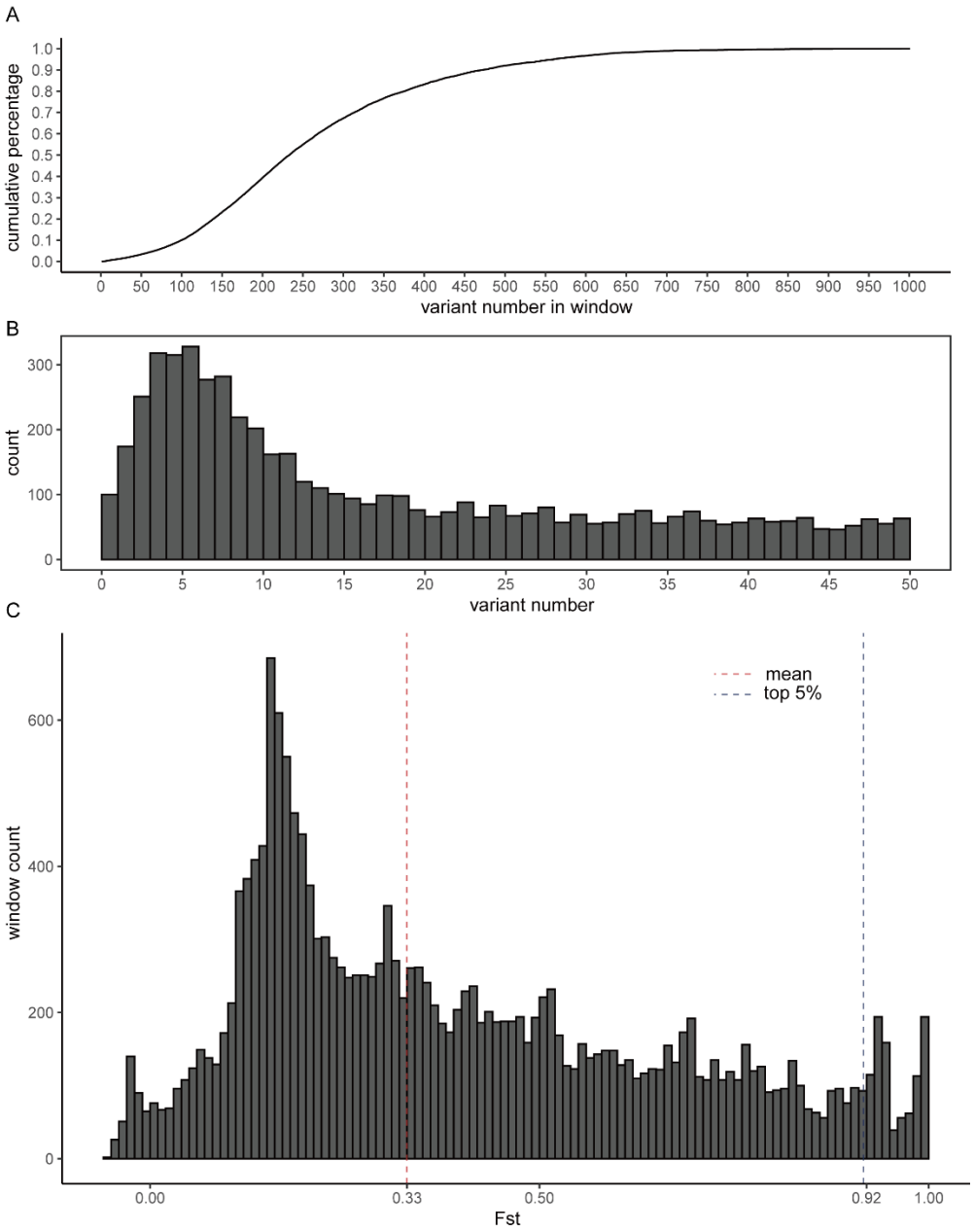

Figure S15. Summary statistic of F_{ST} values in 50-Kb sliding windows for the *B. stacei* African Slope (AS) and European Slope (ES) populations at ECI. **A)** Cumulative curve of SNP count per window size. **B)** Distribution of SNP count in 50-Kb windows, windows containing more than 50 SNPs were omitted from the plot. **C)** Distribution of pairwise F_{ST} values per number of 50-Kb windows along the whole genome of the studied samples.

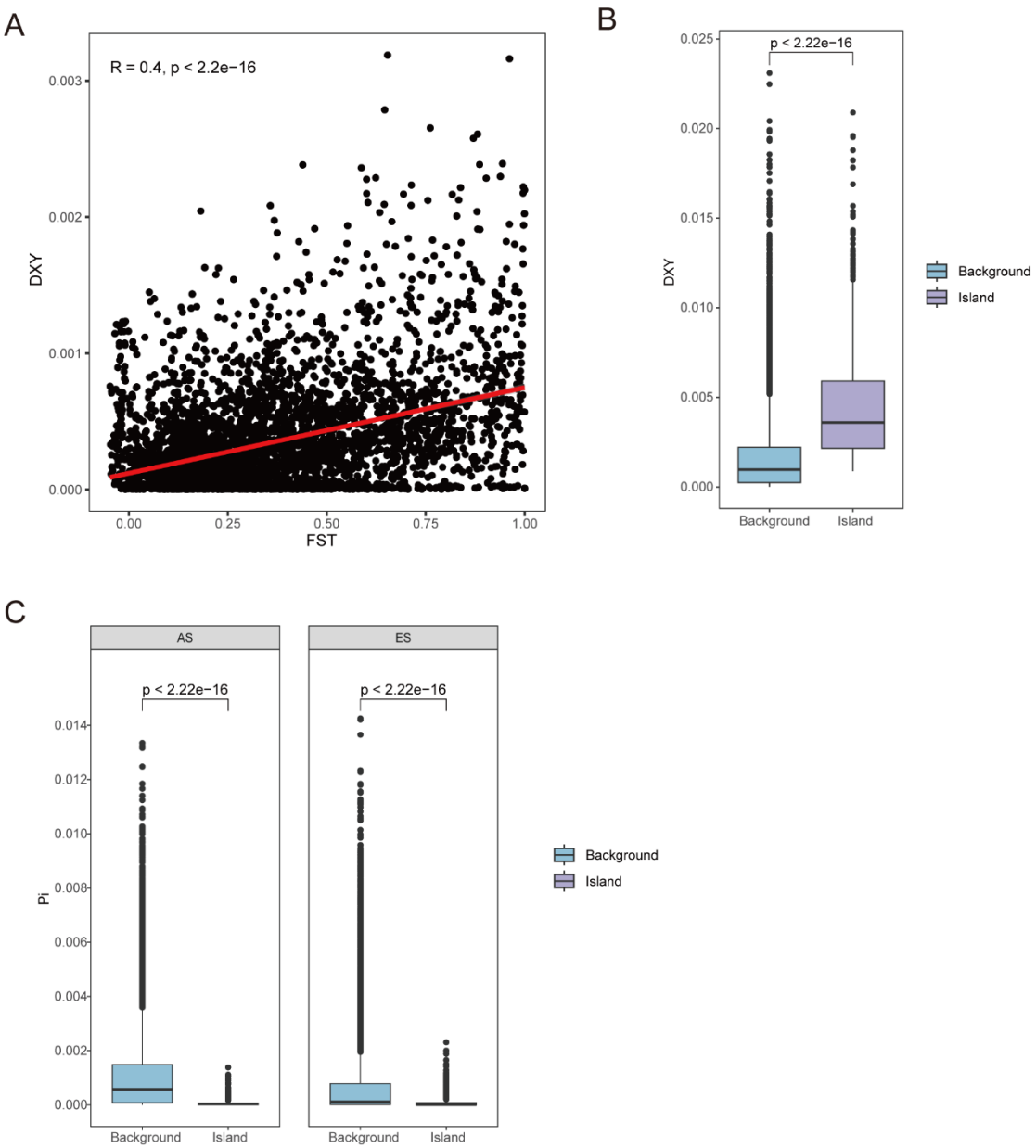

Figure S16. Genetic parameters of whole genomic background and genomic islands in *B. stacei* AS and ES populations from ECI. **A)** Correlation analysis between pairwise F_{ST} fixation values and genetic divergence D_{XY} values (Pearson's correlation test) of whole genomic background. **B)** D_{XY} differences between genomic background and genomic islands (Wilcoxon test); **C)** π differences between genomic background and genomic islands (Wilcoxon test). Colour codes for the respective genomic background and genomic islands are indicated in the charts.

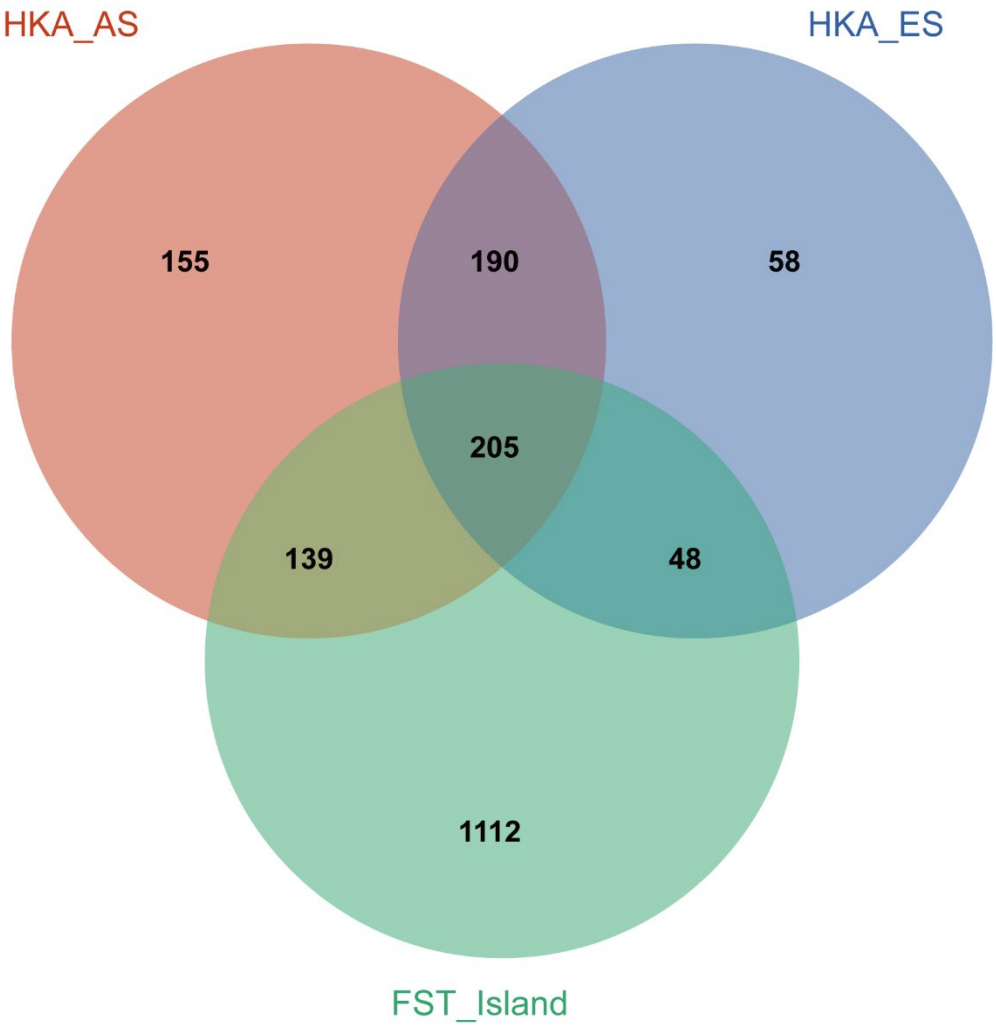

Figure S17. Venn plot of the overlap of candidate adaptive genes showing the overlapping number of genes under recent selection, according to the HKA-tests, from the AS (salmon) and ES (blue) *B. stacei* populations, and those identified through high F_{ST} genomic islands' values from the two compared populations (green).

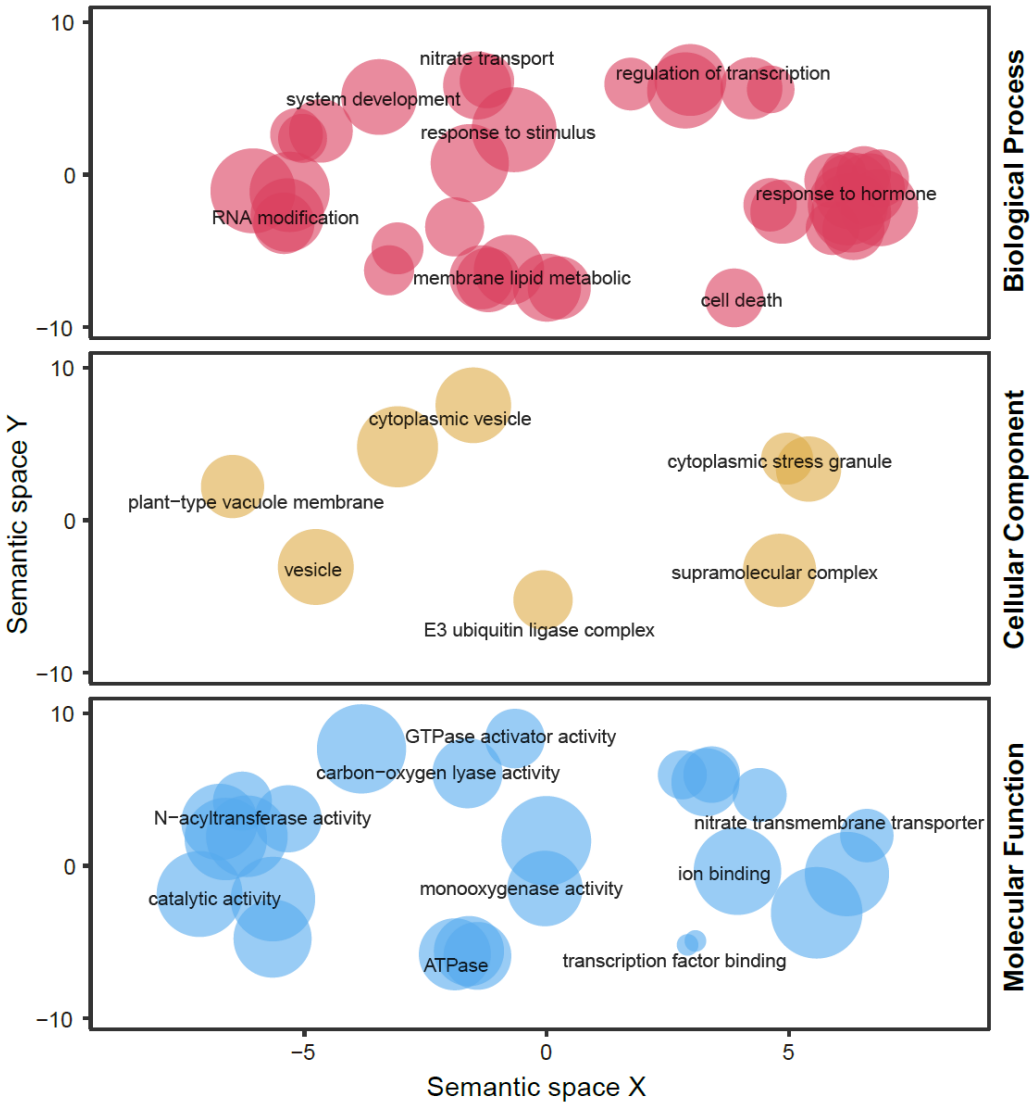

Figure S18. REViGO semantic similarity scatter plots of Gene Ontology (GO) terms for *B. stacei* genes identified by the HKA (Hudson-Kreitman-Aguadé) test as recently selected for Biological processes, Cellular components, and Molecular function. In the semantic spaces, the X and Y axes in the plot represent similarities between GO terms, and the size of the dot indicates gene counts of GO terms.

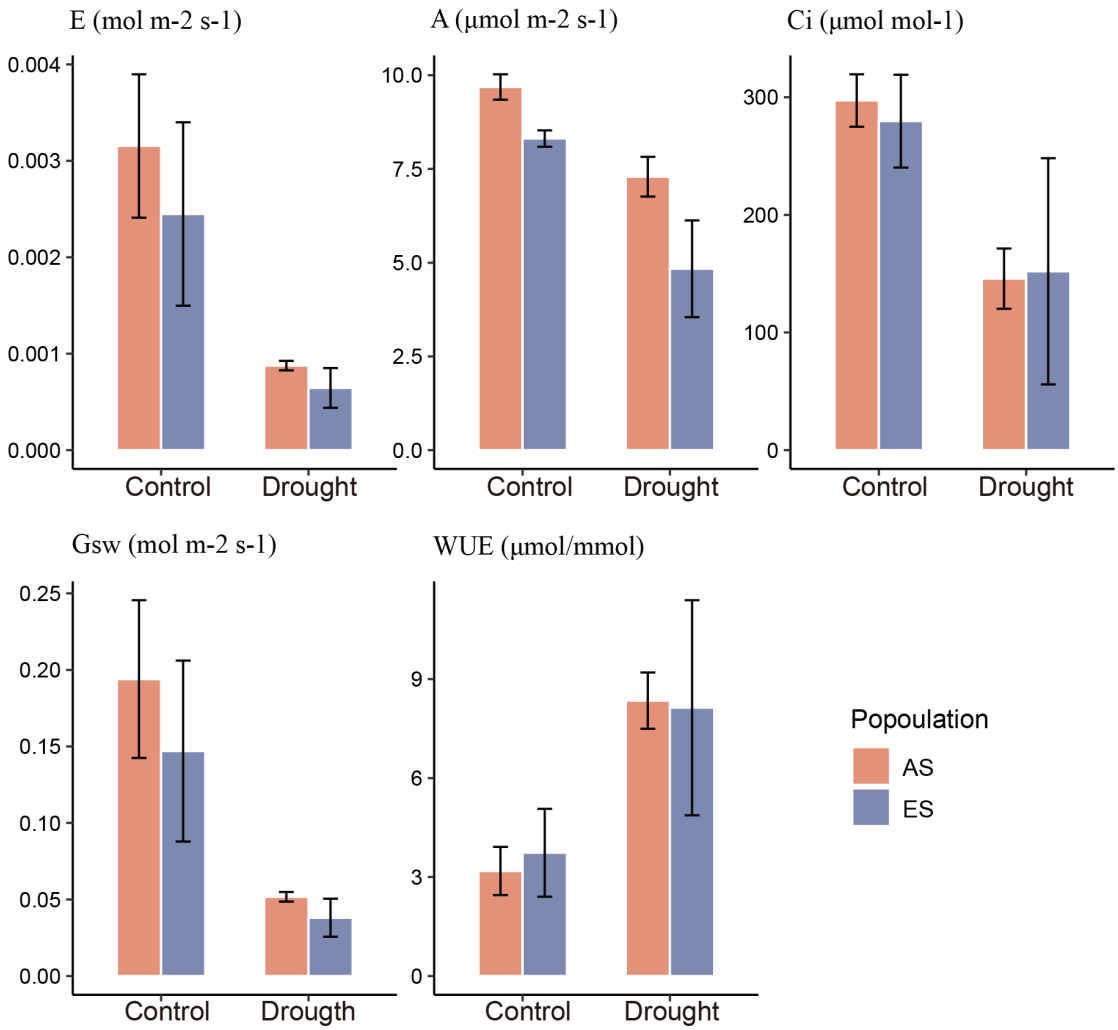

Figure S19. Variation of physiological parameters in samples of the two *Brachypodium stacei* AS and ES populations of ECI under treatment (drought) vs control (well-watered) conditions. E, Transpiration rate; A, Assimilation rate; Ci, intracellular carbon dioxide concentration; Gsw, Stomatal conductance to water vapor; WUE, water use efficiency. Color codes for AS and ES populations are indicated in the chart.

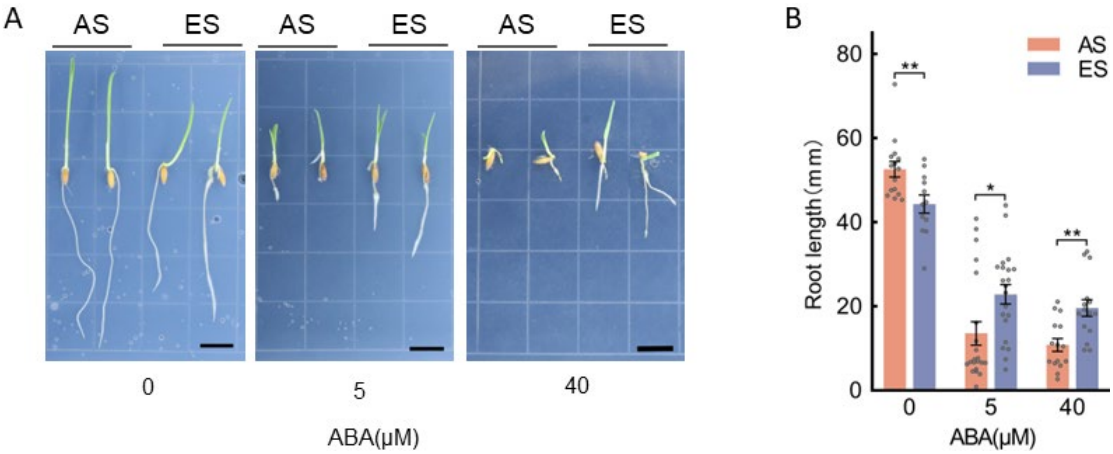

Figure S20. The root growth assay of seedling plants from AS and ES populations subjected to Abcisic acid (ABA) induction. **A)** Root growths under control condition (absence of ABA, left panel), and under different ABA concentrations (5 µM, central panel; 40 µM, right panel); root growth was documented at 7 d post transfer to media (1/2 MS, 1% sucrose) with 0, 5 and 40 µM ABA, respectively. **B)** Statistical analysis of the ABA-inhibited root growth (root lengths of AS and ES ecotypes in each of the three conditions: 0, 5 and 40 µM ABA). Histograms showing mean values and error bars (means ± SE; $n > 20$), * and **: p-value < 0.05 and 0.01 (t-test).

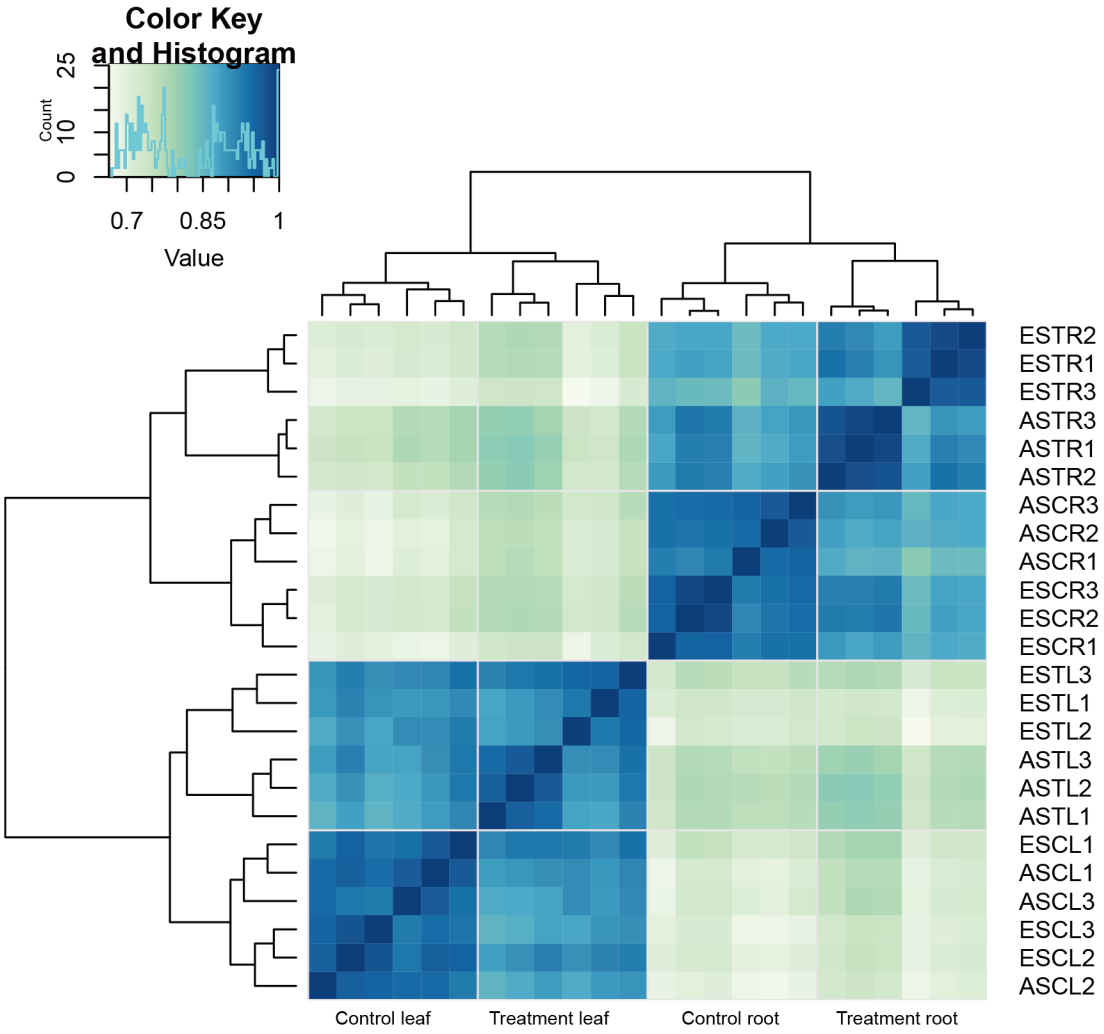

Figure S21. A cluster heat map showing the Pearson correlation of the expressed leaf and root genes between *B. stacei* samples from the African Slope (AS) and European Slope (ES) populations at ECI under drought (treatment) and well-watered (control) conditions. TL: treatment leaf; CL: control leaf; TR: treatment root; CR: control root. Three individual sample replicates from each AS and ES population were used in the expression analysis. Heatmap color code values are indicated in the chart.

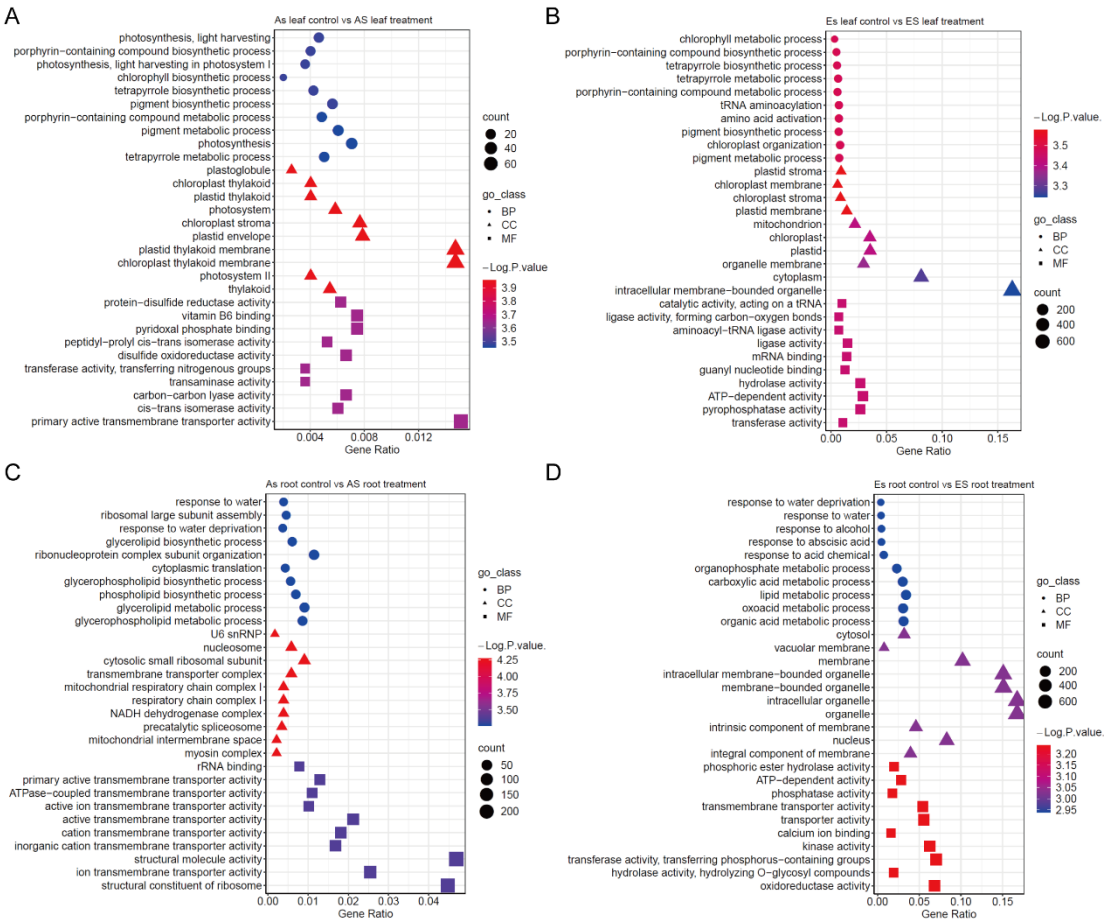

Figure S22. Plots of GO enrichment terms of differentially expressed genes in samples of *B. stacei* from the African Slope (AS) and European Slope (ES) populations at ECI exposed to drought (treatment) *vs* well-watered (control) conditions. Intra-population comparisons, only the top 10 terms are shown. **A**) AS leaf control *vs* AS leaf treatment. **B**) ES leaf control *vs* ES leaf treatment. **C**) AS root control *vs* AS root treatment. **D**) ES root control *vs* ES root treatment. Number of genes (circle size and count), GO class and -logPvalue symbols and color codes are indicated in the respective charts.

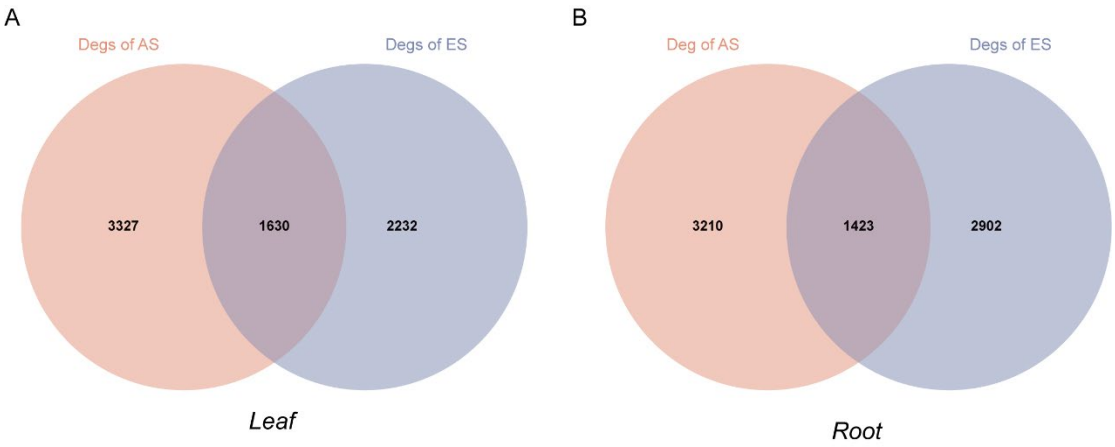

Figure S23. Venn plot of differentially expressed genes (DEGs) in inter-population comparisons of *Brachypodium stacei* samples from the African Slope (AS, salmon) and European Slope (ES, blue) populations at ECI exposed to drought (treatment) *vs* well-watered (control) conditions. **A**) leaf tissue; **B**) root tissue.

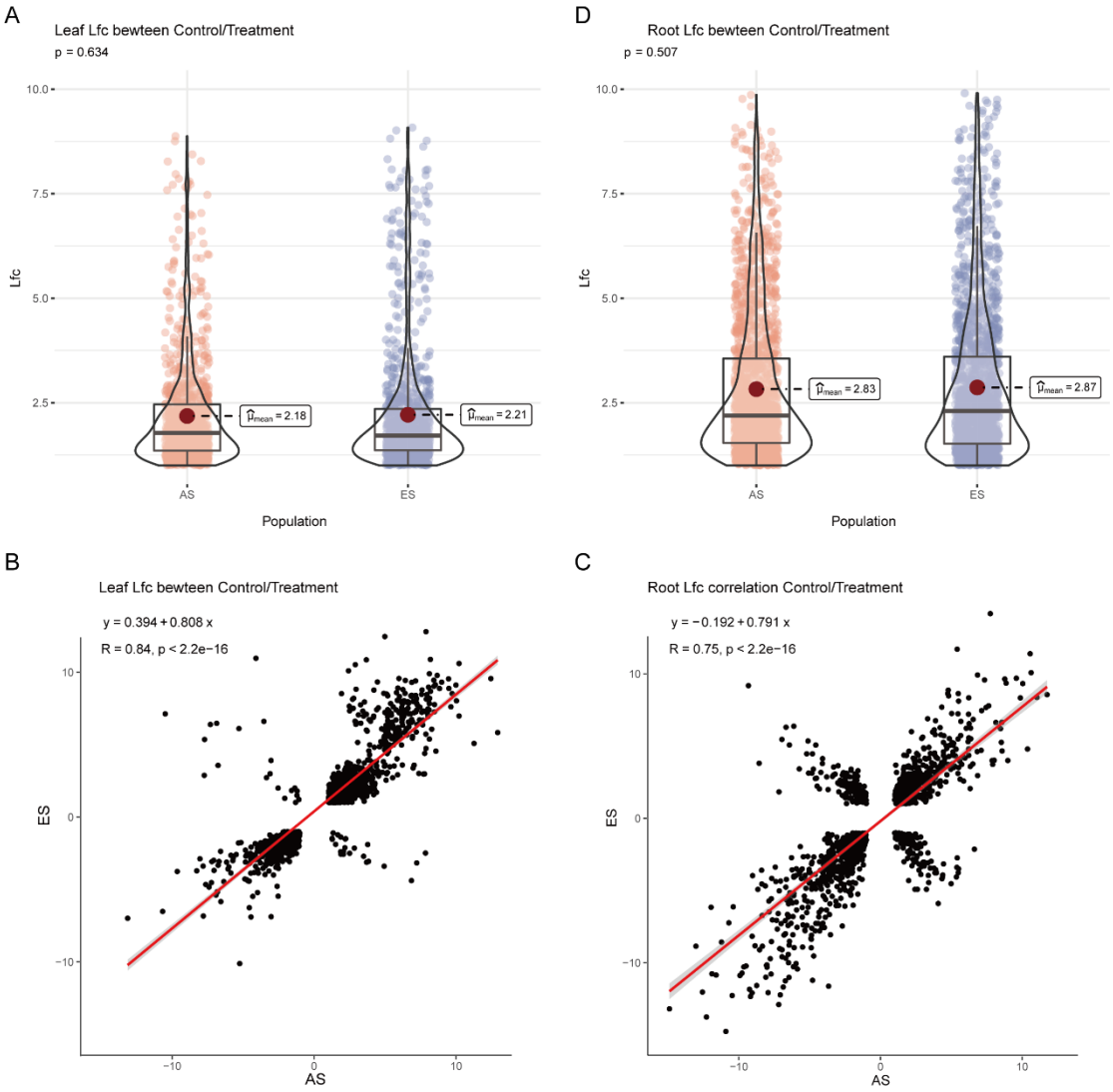

Figure S24. Logarithmic fold change (Lfc) distributions of most up and down regulated genes in intra-population comparisons and Pearson correlation coefficients in inter-population comparisons of common DEGs in *Brachypodium stacei* leaf and root tissues of samples from the African Slope (AS) and the European Slope (ES) populations at ECI under well-watered (control) vs drought (treatment) conditions. Only genes with |Lfc|<10 and Pearson correlation coefficient <|15| are shown. **A**) Lfcs of leaf DEGs in control *vs* treatment conditions in AS and ES samples (t-test). **B**) Lfc of root DEGs in control *vs* treatment conditions in AS and ES samples (t-test). **C**) Pearson correlation coefficient of Lcf leaf DEG values from control *vs* treatment conditions (see subfigure A) between AS and ES samples. **D**) Pearson correlation coefficient of Lcf root DEG values from control *vs* treatment conditions (see subfigure B) between AS and ES samples. Correlation coefficients were high and significant between the two populations for DEGs' Lfc values in both leaf and root tissues.

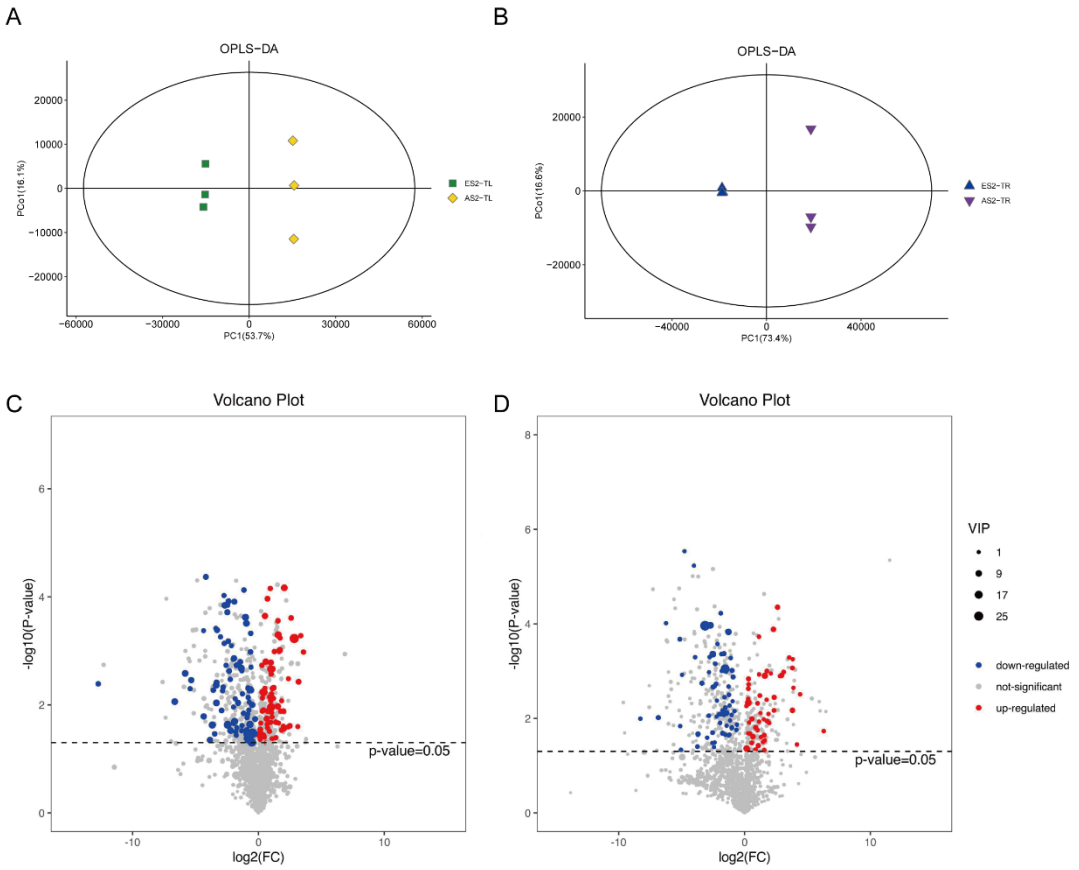

Figure S25. Orthogonal partial least-squares discrimination analysis (OPLS-DA) and Variable importance plot (VIP) of metabolomic profiles of *B. stacei* samples from the African Slope (AS) and European Slope (ES) populations at ECI. **A, B**) Bidimensional OPLS-DA plots of metabolite profiles in leaf **A)** and root **B)** samples under drought conditions from both populations. Symbols and color codes are indicated in the charts. **C, D**) Significantly regulated metabolites (up regulated red; down regulated blue) were identified by |FoldChange| > 1 in leaf (C) and root (D) tissues of the studied populations' samples under drought treatment. TL: (drought) treatment leaf; TR: (drought) treatment root. VIP codes (circle size) and regulation codes (colors) are indicated in the chart.

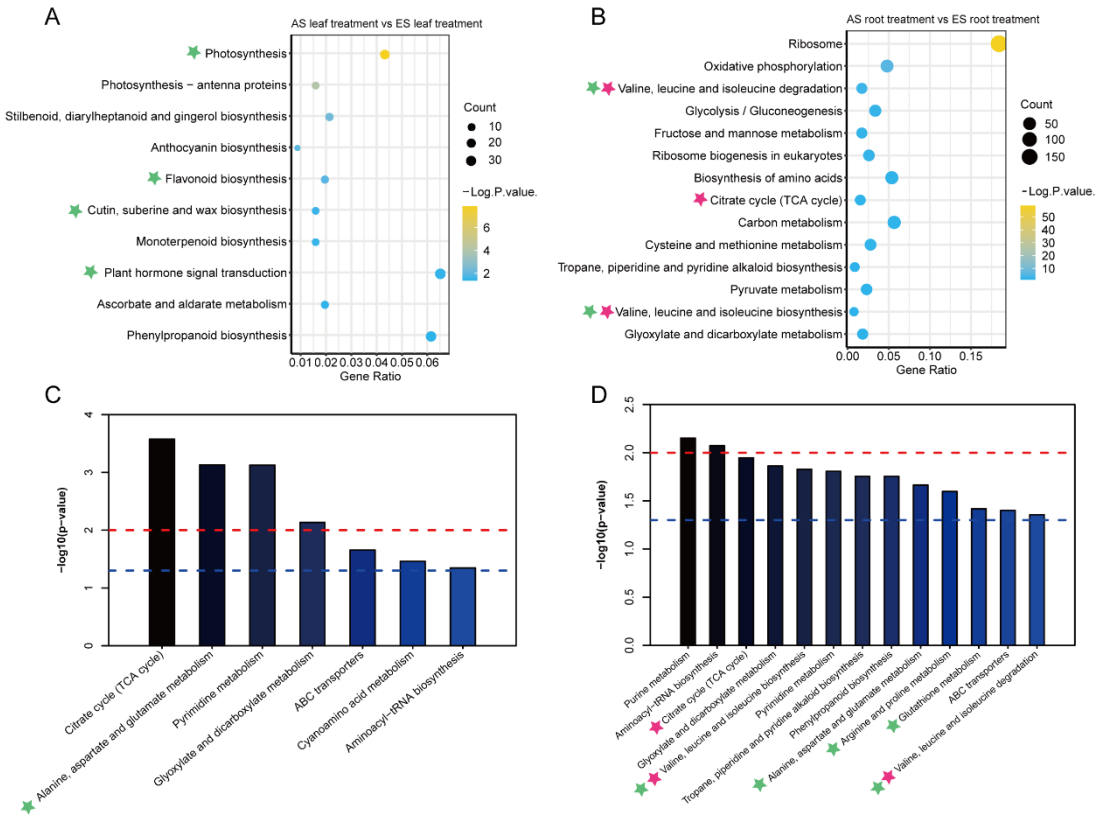

Figure S26. Comparative analysis of enriched transcriptome and metabolome KEGG pathways of *B. stacei* AS and ES population samples from inter-population comparison under drought stress treatment (AS treatment vs ES treatment) for leaf and root tissue samples. Terms involved in drought stress response are marked with a green star and those enriched in both transcriptome and metabolome analysis are marked with a red star. **A, B**) Enriched transcriptomes: (A) DEGs enrichment of AS *vs* ES leaf samples under drought conditions. (B) DEGs enrichment of AS vs ES root samples under drought conditions. Gene count (circle sizes) and -LogPvalue (color codes) are indicated in the respective charts. **C, D**) Enriched metabolomes: (C) Differentially regulated metabolites enrichment of AS vs ES leaf samples under drought conditions. (D) Differentially regulated metabolites enrichment of AS vs ES root samples under drought conditions. Dashed blue and red lines indicate p-values < 0.05 and <0.01, respectively.

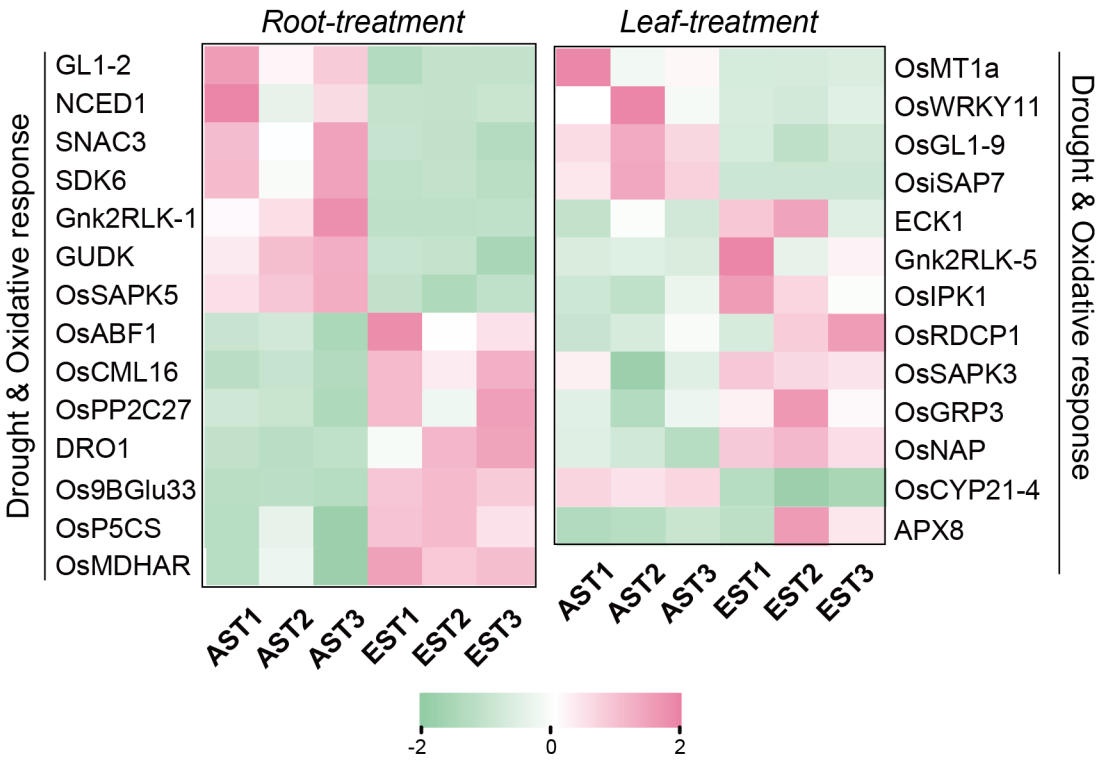

Figure S27. Heatmap of *B. stacei* DEGs for drought response and oxidative response in root and leaf tissues from AS and ES samples under drought (treatment) conditions. AS and ES root and leaf samples consisted of three biological replicates each (T1-T3). Color codes of the heatmap are indicated in the chart.

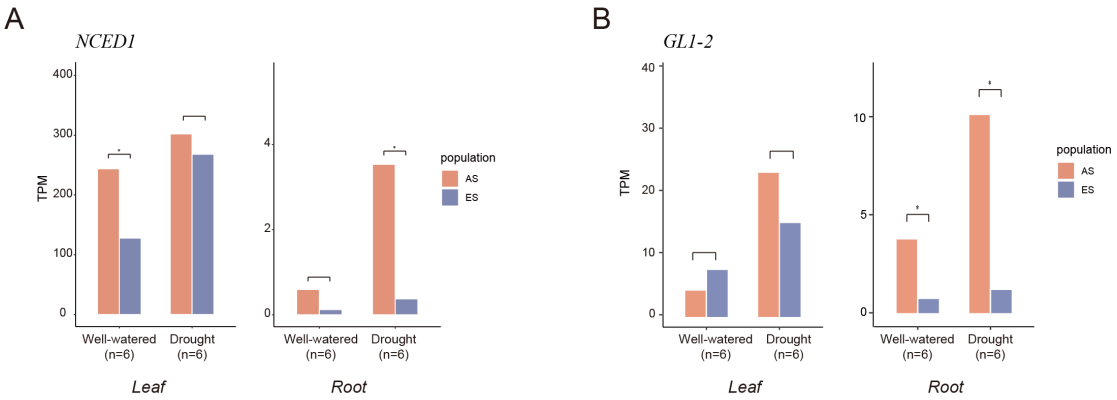

Figure S28. Expression levels of *NCED1* and *GL1-2* in transcripts per million (TPM) in leaf and root samples of *B. stacei* samples from the AS and ES populations under well-watered (control) vs drought stress (treatment) conditions; **A)** *NCED1*; **B)** *GL1-2*. * p-value < 0.05 (Wilcoxon rank sum test).

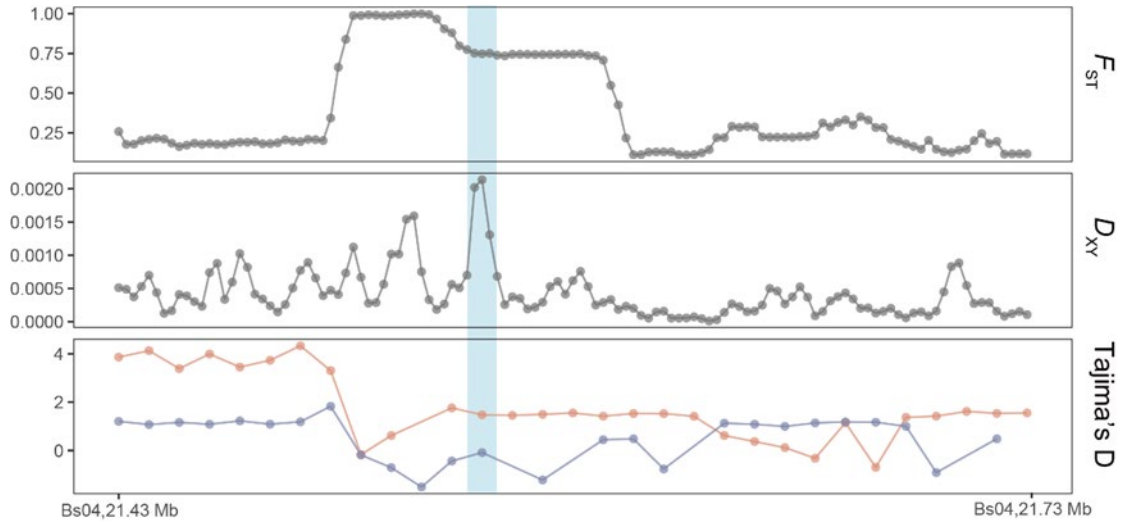

Figure S29. Chromosomal location of a Glossy1 (GL1) B. stacei orthologous GL1-2 gene in chromosome Bs04 of the Bsta-ECI local reference genome. Pairwise FST, DXY and Tajima's D values from GL1-2 sequences in the sequenced African Slope (AS) and European Slope (ES) individuals from Evolution Canyon I.

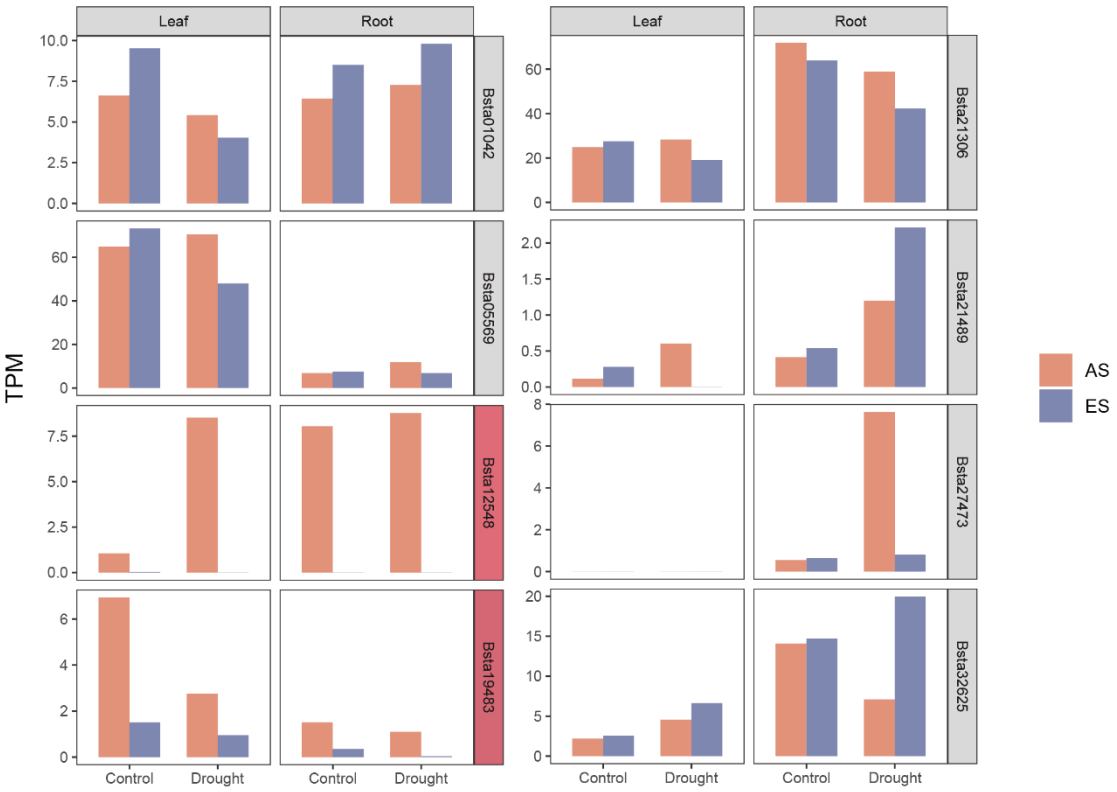

Figure S30. The expression level of 9 genes nearby TIPs in the different *B. stacei* AS and ES population samples' tissues and conditions. The genes identified as DEG in all inter-population comparisons were marked as red. AS and ES color codes are indicated in the chart.

Table S1. Characteristics of the genome sequencing data for the Bsta-ECI reference genome from Evolution Canyon I, Israel (ECI). (See Dataset S1).

Method	Library type	Reads number	Data size (Gb)	Mean read length (bp)	Read N50 (bp)
Illumina (clean data)	Paired	115,050,912	17.24	150	-
Pacbio Hifi	Single	1,639,829	25.01	15253	15357
Hi-C (clean data)	Paired	1,011,713,862	151.75	150	-

Table S2. Comparison of genome assembly parameters of Bsta-ECI and the *B. stacei* ABR114 reference genome generated by the Joint Genome Institute (JGI; Phytozome https://phytozome-next.jgi.doe.gov/; *B. stacei* ABR114 v. 1.1).

Genome feature	***B. stacei* (ECI)**	***B. stacei* (ABR114 v.1.1)**
Total length of contigs (Mb)	256.71	-
Total length of assemblies (Mb)	248.99	234.14
Gap number	30	3,132
Number of contigs	79	3,244
Contig N50 (Mb)	10.19	0.23
Number of scaffolds	10	10
Scaffold N50 (Mb)	24.13	23.06
LAI	10.46	2.06
GC content (%)	45.53	44.71
Percentage of anchoring (%)	96.99	-

Table S3. Summary of Hi-C mapping parameters of the Bsta-ECI genome from Evolution Canyon I.

	***B. stacei* Bsta-ECI**	
	Read number	**Percentage (%)**
Total read pairs	502,167,104	100
Mapped read pairs	425,121,335	84.66
Uniquely mapped read pairs	313,511,156	62.43
Valid interaction read pairs	258,389,405	51.45

Table S4. Chromosome level comparisons of the assembled Bsta-ECI to the reference genome *B. stacei* ABR114 v 1.1 (Phytozome; https://phytozome-next.jgi.doe.gov/).

	Bsta-ECI		ABR114 v. 1.1	
	Chromosome Length (bp)	Gap Number	Chromosome Length (bp)	Gap Number
Bs01	32,090,747	4	29,836,869	313
Bs02	29,360,157	4	27,560,134	242
Bs03	27,566,100	0	25,035,503	339
Bs04	25,728,233	2	24,459,551	277
Bs05	24,136,534	3	22,826,389	279
Bs06	23,726,300	3	21,437,706	416
Bs07	22,169,934	7	20,606,078	310
Bs08	22,261,700	3	20,433,684	355
Bs09	21,987,753	1	20,329,407	271
Bs10	19,969,809	2	18,553,268	314

Table S5. Summary of genome assembly validation and statistics of the newly assembled Bsta-ECI genome and the reference genome *B. stacei* ABR114 v.1.1

	B. stacei Bsta-ECI	*B. stacei* ABR114 v1.1
BUSCOs (%)	98.6	98.4
Complete and single-copy BUSCOs (%)	97.3	97.1
Complete and duplicated BUSCOs (%)	1.3	1.3
Fragmented BUSCOs (%)	0.4	0.5
Missing BUSCOs (%)	1.0	1.1
Mapping statistics		
Fraction of Mapped Illumina Data (%)	99.51	97.24
Fraction of Properly read pairs Mapped (%)	99.16	96.41
Regions of Coverage > 0x (%)	99.74	98.85
Regions of Coverage ≥ 10x (%)	99.63	98.82
K-mer		
Base call accuracy (QV)	39.79	40.18
Completeness	99.10	96.16

Table S6. Prediction of repetitive elements in the assembled Bsta-ECI genome from Evolution Canyon I.

Type	Repeat Size (bp)	Percent of genome (%)
TRF[1]	18,301,727	7.13
RepeatMasker	50,048,364	19.50
RepeatProteinMask	16,203,285	6.51
De novo	97,141,916	37.84
Total	103,467,454	40.31

1: TRF: Tandem Repeat Finder.

Table S7. Summary of repetitive elements in the Bsta-ECI genome and in the *B. stacei* ABR114 v 1.1 reference genome

Repetitive element type	Length(bp) of genome		Percentage (%) of Repeat type		Percentage (%) of genome covered by repeat type	
	Bsta-ECI	ABR114	Bsta-ECI	ABR114	Bsta-ECI	ABR114
SINE	611,446	436,828	0.591	0.57	0.238	0.19
LINE	7,583,328	6,785,199	7.329	8.86	2.954	2.90
L1	7,492,395	6,756,750	7.241	8.45	2.919	3.21
Other	210,321	29,953	0.2	0.04	0.08	0.01
LTR	46,478,386	35,045,130	44.921	45.77	18.106	14.97
Copia	16,894,953	10,043,204	16.329	13.12	6.581	4.29
Gypsy	29,297,286	23,934,981	28.315	31.26	11.413	10.22
Other	4,543,886	1,496,957	4.392	1.96	1.77	0.64
DNA	18,532,599	16,158,902	17.912	21.1	7.219	6.90
CMC-EnSpm	3,537,181	2,786,276	3.419	3.64	1.378	1.19
hAT-Ac	3,979,100	897,819	3.846	1.17	1.55	0.38
hAT-Tip100	2,780,739	571,236	2.688	0.75	1.083	0.24
MuDR	1,020,941	3,313,096	0.987	4.33	0.398	1.42
PIF-Harbinger	443,859	2,724,305	0.429	2.44	0.173	1.01
Other	7,339,227	6,325,645	7.093	8.26	2.859	2.70
Transposable elements(TEs) total	73,205,759	58,426,059	70.75	76.3	28.52	24.95
Satellites	126,999	14,724	0.123	0.02	0.049	0.01
Simple_repeats	19,846,896	5,427,411	19.182	7.09	7.731	2.32
Unclassified	29,067,671	16,443,949	28.09	21.47	10.38	7.02
Total repeats	103,467,454	76,575,937	100	100	40.31	32.70

Table S8. Gene predictions and statistics in the Bsta-ECI genome, compared to those of the reference genomes of *B. stacei* ABR114 v. 1.1, , and *B. distachyon* Bdis-Bd21 generated by the Joint Genome Institute (Phytozome, https://phytozome-next.jgi.doe.gov/), and of *Oryza sativa* subsp. *japonica*, generated by National Center for Biotechnology Information (NCBI_Assembly: GCF_001433935.1).

Species	Gene Predicted	Average Gene Length (bp)	Average CDS Length (bp)	Average Exons per Gene	Average Exon Length (bp)	Average Intron Length (bp)	Complete BUSCO (%)
Bsta-ECI	32,951	2528.92	1159.56	4.59	252.13	380.47	99.1
Bsta-ABR114	29,898	3334.01	1197.6	4.76	251.26	393.98	99.2
Bdis-Bd21	34,310	3373.23	1117.39	4.39	254.36	403.07	99.5
O. sativa japonica	37,869	2986.76	982.71	3.74	262.43	426.67	99.7

Table S9. Functional annotation of the predicted genes for Bsta-ECI genome

	Database	*B. stacei* Bsta-ECI	
		Number	**Percentages (%)**
Total		32,951	100
Annotated	NR	30,315	92.00
	InterPro	29,646	89.96
	Swissport	20,785	63.07
	TrEMBL	30,895	93.76
	EggNOG	27,760	84.24
Unannotated		872	2.64

Table S10. Transcription factor (TF) prediction in Bsta-ECI genome.

TF family	*B. stacei* **Bsta-ECI**
Alfin-like	9
AP2/ERF-AP2	25
AP2/ERF-ERF	124
AP2/ERF-RAV	4
B3	52
B3-ARF	26
BBR-BPC	3
BES1	7
bHLH	127
bZIP	85
C2C2-CO-like	9
C2C2-Dof	29
C2C2-GATA	29
C2C2-LSD	6
C2C2-YABBY	8
C2H2	113
C3H	56
CAMTA	7
CPP	9
CSD	4
DBB	4
DBP	6
DDT	6
E2F-DP	10
EIL	6
FAR1	88
GARP-ARR-B	8
GARP-G2-like	49
GeBP	15
GRAS	58
GRF	12
HB-BELL	14
HB-HD-ZIP	38
HB-KNOX	11
HB-other	12
HB-PHD	3
HB-WOX	13
HRT	1
HSF	25
LFY	1
LIM	6
LOB	26

MADS-MIKC	34
MADS-M-type	43
MYB	122
MYB-related	65
NAC	122
NF-X1	2
NF-YA	7
NF-YB	13
NF-YC	13
OFP	30
PLATZ	15
RWP-RK	16
S1Fa-like	1
SBP	17
SRS	6
STAT	1
TCP	22
Tify	15
Trihelix	28
TUB	12
ULT	1
VOZ	2
Whirly	2
WRKY	84
zf-HD	21
All	1,838
Percentage of total gene predictions	5.5%

Table S11. Summary of genomic variants obtained in the 41 *B. stacei* individuals studied in Evolution Canyon I.

Chromosome	Length	SNP number	SNP rate(%)	INDEL number	INDEL rate (%)
Bs01	32,090,747	96,922	0.30	28375	0.09
Bs02	29,360,157	65,230	0.22	20286	0.07
Bs03	27,566,100	95,553	0.34	25950	0.09
Bs04	25,728,233	74,013	0.29	23583	0.09
Bs05	24,136,534	56,799	0.23	19288	0.08
Bs06	23,726,300	63,596	0.27	19343	0.08
Bs07	22,169,934	86,554	0.40	26363	0.12
Bs08	22,261,700	77,838	0.35	23108	0.10
Bs09	21,987,753	30,580	0.14	11920	0.05
Bs10	19,969,809	75,266	0.38	22105	0.11
All	248,997,267	722,351	0.29	220321	0.09

Table S12. Summary of Structure Variants (SVs) detected among the 41 *B. stacei* resequenced genomes of Evolution Canyon I (Del: deletion; Dup: duplication; Inv: inversion; Ins: insertion; Tra: translocation; UNK: unknow) relative to the Bsta-ECI reference genome.

SVs Len	Del	Dup	Inv	Ins	Tra	UNK
0-50bp	0	0	2	0	5954	0
50-100bp	1352	97	32	1014	0	0
100-1000bp	2023	156	90	560	0	0
1000-10000bp	1016	242	104	0	0	0
10000+bp	283	232	654	0	0	0

Table S13. Comparison of performances of seven alternative models of inference of evolutionary demographic history of AS and ES populations in ECI conducted with *∂a∂i* using forward simulation and residuals analysis. The best asym_mig_twoepoch model was selected based on likelihood and AIC (Akaike Information Criterion) values. Each model was run 100 rounds using the dadi_pipeline (see methods); best likelihood and AIC values for each model are shown.

models	Best likelihood	Best AIC
asym_mig_twoepoch	-22639.63	45295.26
no_mig	-34200.1	68406.2
sym_mig	-25200.15	50408.3
asmy_mig	-23854.14	47718.28
sym_mig_twoepoch	-24925.77	49863.54
ancestor_divergence	-26997.49	54004.98
second_contact	-24060.63	48131.26

Table S14. Parameters of the best asym_mig_twoepoch evolutionary demographic history model for the AS and ES populations estimated from *∂a∂i*. NA, real population size of MRCA; nu1 and nu2, real population sizes of AS and ES populations; m12a and m21a, migration rates from AS to ES and from ES to AS in time slice T1; m12b and m21b, migration rates from AS to ES and from ES to AS in time slice T2; T1, initial time of epoch T1 (to T2); T2, initial time of epoch T2 (to present).

Parameter	∂a∂i output	Transference to readable
NA	1	13421.18
nu1	0.3357	4505.49
nu2	0.1978	2654.71
m12a	3.1526	0.00011
m21a	1.7528	$6.530\ e^{-5}$
m12b	1.542	$5.744\ e^{-5}$
m21b	1.3721	$5.111\ e^{-5}$
T1	0.1131	10167.89
T2	0.3788	3035.87

Table S15. Distribution of 'genomic island' located on each chromosome.

Chr	Outlier windows distributed over one chromosome	windows distributed over one chromosome	Island length >100 kb (long)	Island length 10-100 kb (short)	P-value (Fisher.test)
Bs01	165	2,977	4	41	0.313
Bs02	118	2,648	4	27	0.073
Bs03	24	2,603	0	12	1
Bs04	72	2,388	0	30	0.622
Bs05	58	2,169	1	13	0.498
Bs06	126	2,214	2	36	0.693
Bs07	64	2,118	1	19	1
Bs08	73	2,146	0	51	0.231
Bs09	98	1,770	1	30	1
Bs10	9	1,899	0	5	1
Total	807	22,932	13	271	

Table S16. Detection of a putative *B. stacei* ortholog to rice *OsNCED1,* a key regulatory gene in ABA biosynthesis, through reciprocal Blast searches. Only the top 5 hits of each search are shown. The best match corresponds to *Bsta14910* (84.3% identity).

query id	subject id	Percentage of identity	evalue	bitscore
Os02g0704000	Bsta14910	84.3	0	917
Os02g0704000	Bsta06087	42.3	2.04E-122	375
Os02g0704000	Bsta21788	40.8	1.06E-113	353
Os02g0704000	Bsta32571	39.7	1.60E-106	335
Os02g0704000	Bsta32687	37.6	1.36E-101	319
Bsta14910	Os02g0704000	84.3	0	918
Bsta14910	Os12g0435200	66.5	2.38E-255	714
Bsta14910	Os03g0645900	42.6	1.44E-117	363
Bsta14910	Os07g0154100	42.3	3.58E-114	353
Bsta14910	Os12g0640600	37.9	2.14E-94	300

Table S17. Location of 16 SNPs in the 5'-upstream regulatory region of the *Bsta14910* gene**,** a putative ortholog of *OsNCED1*, separating the AS and ES population samples' sequences, that show different haplotypes between them. AS-ES allelic frequencies of polymorphic *Bsta14910* positions were compared relative to the *Bsta14910* sequence of the Bsta-ECI reference genome.

Bsta-ECI chromosome	Bsta-ECI position	AS representative seq frequence	AS allelic frequence	ES representative seq frequence	ES allelic frequence
Bs04	5823937	T:0.791	G:0.208	T:0	G:1
Bs04	5824072	C:1	T:0	C:0	T:1
Bs04	5824375	A:0.8	T:0.2	A:0	T:1
Bs04	5824376	C:0.8	G:0.2	C:0	G:1
Bs04	5824706	C:1	T:0	C:0	T:1
Bs04	5824961	G:0.8	C:0.2	G:0	C:1
Bs04	5825020	A:0.8	G:0.2	A:0	G:1
Bs04	5825029	C:0.8	A:0.2	C:1	A:0
Bs04	5825092	A:0.8	G:0.2	A:1	G:0
Bs04	5825111	T:0.8	C:0.2	T:0	C:1
Bs04	5825184	A:0.8	G:0.2	A:0	G:1
Bs04	5825199	A:0.8	T:0.2	A:0	T:1
Bs04	5825222	T:0.8	C:0.2	T:0	C:1
Bs04	5825237	A:1	T:0	A:0	T:1
Bs04	5825415	A:0.8	G:0.2	A:0	G:1
Bs04	5825601	G:0.8	A:0.2	G:0	A:1

SI References

1. E. Nevo, "Evolution Canyon," a potential microscale monitor of global warming across life. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2960–2965 (2012).
2. M. Kokot, M. Dlugosz, S. Deorowicz, KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).
3. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11** (2020).
4. S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
5. N. C. Durand, *et al.*, Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Syst.* **3**, 95–98 (2016).
6. O. Dudchenko, *et al.*, De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science (80-.).* **356**, 92–95 (2017).
7. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury : reference-free quality , completeness , and phasing assessment for genome assemblies. 1–27 (2020).
8. V. Md, S. Misra, H. Li, S. Aluru, Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *Proc. 2019 IEEE 33rd Int. Parallel Distrib. Process. Symp. IPDPS 2019*, 314–324 (2019).
9. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
10. K. R. Bradnam, *et al.*, Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 1–31 (2013).
11. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
12. S. Ou, N. Jiang, LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
13. M. Zhu, *et al.*, Multi-omics reveal differentiation and maintenance of dimorphic flowers in an alpine plant on the Qinghai-Tibet Plateau. *Mol. Ecol.* **32**, 1411–1424 (2023).
14. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
15. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21**, 351–358 (2005).

16. N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, 1–14 (2004).
17. J. Jurka, *et al.*, Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
18. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
19. M. Stanke, *et al.*, AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, 435–439 (2006).
20. S. P. Gordon, *et al.*, Gradual polyploid genome evolution revealed by pan-genomic analysis of Brachypodium hybridum and its diploid progenitors. *Nat. Commun.* **11**, 1–16 (2020).
21. J. Keilwagen, F. Hartung, J. Grau, GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177 (2019).
22. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
23. M. Pertea, *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
24. B. J. Haas, *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
25. B. J. Haas, *et al.*, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
26. B. J. Haas, *et al.*, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 1–22 (2008).
27. P. Jones, *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
28. A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
29. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, 61–65 (2007).
30. J. Huerta-Cepas, *et al.*, EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

31. A. Conesa, *et al.*, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
32. Y. Zheng, *et al.*, iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
33. L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
34. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
35. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLoS Genet.* **2**, 2074–2093 (2006).
36. A. Manichaikul, *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
37. Felsenstein, J. PHYLIP-Phylogeny inference package (version 3.2). Cladistics 5, 164–166 (1989).
38. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
39. C. Zhang, S. S. Dong, J. Y. Xu, W. M. He, T. L. Yang, PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
40. V. Shiposha, P. Catalán, M. Olonova, I. Marques, Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the iberian peninsula and into the islands. *PeerJ* **2016** (2016).
41. C. C. Chang, *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).
42. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5** (2009).
43. D. M. Portik, *et al.*, Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* **26**, 5245–5263 (2017).
44. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).

45. T. Ma, *et al.*, Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E236–E243 (2017).
46. J. Zhang, S. Zhang, Z. Zheng, Z. Lu, Y. Yang, Genomic divergence between two sister Ostrya species through linked selection and recombination. *Ecol. Evol.* **12**, e9611 (2022).
47. R. R. Hudson, M. Kreitman, M. Aguadé, A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
48. S. Liu, *et al.*, Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785–794 (2014).
49. J. J. Jin, *et al.*, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. **21**, 1–31 (2020).
50. X. J. Qu, M. J. Moore, D. Z. Li, T. S. Yi, PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**, 1–12 (2019).
51. K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
52. J. Rozas, *et al.*, DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
53. J. W. Leigh, D. Bryant, POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
54. A. Conesa, *et al.*, A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 1–19 (2016).
55. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
56. D. V. Klopfenstein, *et al.*, GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 1–17 (2018).
57. T. Wu, *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*. **2**, 100141 (2021).