RESEARCH

Open Access

Genomic determinants, architecture, and constraints in drought-related traits in *Corymbia calophylla*

Collin W. Ahrens[1,2*], Kevin Murray[3], Richard A. Mazanec[4], Scott Ferguson[3], Ashley Jones[3], David T. Tissue[1], Margaret Byrne[4], Justin O. Borevitz[3] and Paul D. Rymer[1]

Abstract

Background Drought adaptation is critical to many tree species persisting under climate change, however our knowledge of the genetic basis for trees to adapt to drought is limited. This knowledge gap impedes our fundamental understanding of drought response and application to forest production and conservation. To improve our understanding of the genomic determinants, architecture, and trait constraints, we assembled a reference genome and detected ~ 6.5 M variants in 432 phenotyped individuals for the foundational tree *Corymbia calophylla*.

Results We found 273 genomic variants determining traits with moderate heritability ($h^2_{SNP} = 0.26–0.64$). Significant variants were predominantly in gene regulatory elements distributed among several haplotype blocks across all chromosomes. Furthermore, traits were constrained by frequent epistatic and pleiotropic interactions.

Conclusions Our results on the genetic basis for drought traits in *Corymbia calophylla* have several implications for the ability to adapt to climate change: (1) drought related traits are controlled by complex genomic architectures with large haplotypes, epistatic, and pleiotropic interactions; (2) the most significant variants determining drought related traits occurred in regulatory regions; and (3) models incorporating epistatic interactions increase trait predictions. Our findings indicate that despite moderate heritability drought traits are likely constrained by complex genomic architecture potentially limiting trees response to climate change.

Keywords Eucalyptus, Epistasis, Pleiotropy, Genome wide association study (GWAS), Water use efficiency, Heritability

Introduction

Climate change is increasing the intensity and frequency of droughts worldwide [1], pushing trees to their physiological limits, and in some cases to the point of failure, resulting in forest dieback [2]. A species's ability to tolerate drought is likely determined by complex genome characteristics, including base pair changes [3, 4], large rearrangements [5], and/or interactions between genes [6, 7]. Understanding the genetic mechanisms that control drought related traits can lead to better predictions of drought tolerance, increasing our success in managing natural and planted forests under climate change induced drought.

*Correspondence:
Collin W. Ahrens
collinwahrens@gmail.com
[1] Hawkesbury Institute for the Environment, Western Sydney University, Richmond, NSW 2753, Australia
[2] Cesar Australia, Brunswick, VIC 3058, Australia
[3] Research School of Biology, Australian National University, Canberra, ACT 2600, Australia
[4] Biodiversity and Conservation Science, Western Australian Department of Biodiversity, Conservation and Attractions, Kensington, WA 6151, Australia

Droughts are major selective forces [8, 9], however it is generally unknown how much variation of drought tolerant traits in trees are genetically controlled. There are many traits that provide tolerance to drought conditions. One drought trait that stands out is carbon isotope discrimination or the ratio between C13 and C12 (δ^{13}C). Isotope discrimination is important because it is based on Rubisco's preference for light carbon (C12), and plants with a higher proportion of C13 are generally more drought tolerant [10]. Further, δ^{13}C has been strongly correlated to stomatal conductance [11] and water use efficiency [12]. Most studies that link genotype to drought tolerance (as δ^{13}C) have been performed on agriculturally important species, such as a study that shows some evidence of genetic control of δ^{13}C in soybeans [13]. However, a recent study on an ecologically important species identified 78 and 6 drought tolerant variants related to δ^{13}C in coast redwoods and giant sequoias, respectively [14], and another study used a few thousand single nucleotide polymorphisms (SNPs) for Q_{ST}–F_{ST} comparisons to identify selection occurring for δ^{13}C in *Pinus pinaster* [15]. A second often used trait that is associated with drought tolerance is specific leaf area (SLA), which is the leaf surface area per unit of dry biomass [16, 17]. A study on *Populus trichocarpa* found two SNPs associated with this trait [18]. However, other studies suggest that SLA is highly plastic and largely not heritable [19, 20]. A third trait used to quantify the effects of drought is the normalized difference vegetation index (NDVI) which measures chlorophyll reflectance [21]. A study on maize found nine potential adaptive SNPs controlling NDVI [22]. Natural selection acts directly upon expressed phenotypes [23], which are controlled by additive and non-additive genetic variation [24]. Most studies linking genotype and phenotype using GWAS focus on quantifying the additive genetic variation controlling the trait of interest [24, 25]. However, by explicitly understanding the contribution of both additive and non-additive genetic variation to drought tolerant traits, we could improve predictions on how well populations can respond to novel drought conditions.

It is inherently difficult to quantify the genotypic effect on physiological traits when measuring plants in situ because the variation could be due to environment and not genotype [26]. Growing related individuals from many populations in a common garden minimises the environmental variance [27] resulting in the phenotypic variance being the product of the genotypic differences. This allows for the estimation of trait heritability, which can be interpreted as part of the trait's 'evolvability' or evolutionary potential [28]. Common gardens are also an important resource to use in conjunction with genome-wide association studies (GWAS), which explicitly evaluates each correlation between SNP and trait using mixed effects linear models. GWAS studies have been deployed for a vast number of species to understand the genetic determination of trait variation and gene discovery [29–31]. Therefore, GWAS is a powerful technique that allows for the identification of within and among population standing genetic variation related to complex traits.

Estimates of trait heritability can be determined based on genetic variants (i.e., SNPs) contributing to phenotypic variation in GWAS analyses using an additive genetic framework. However, heritability from GWAS analyses is unable to account for non-additive factors that could contribute to the heritability of a trait. Missing heritability in complex traits may be due to non-additive genetic variation such as gene–gene interactions (epistasis [32]) and gene-trait interactions (pleiotropy [33]). Accounting for epistatic and pleiotropic interactions that contribute to the heritability of quantitative traits can improve model predictions [33, 34]. Epistatic effects could enhance or remove the effect of a gene on the trait depending on the presence of interacting genes in the genetic background. The widespread presence of epistasis could limit our ability to identify all the heritable variation but could constrain or boost adaptation of traits. Another biological process that could potentially constrain traits is pleiotropy because causal genes could be potentially interacting with multiple traits [35]. Pleiotropy could result in an antagonistic behaviour where one gene positively affects one trait while negatively affecting a second trait [36]. Together, epistasis and pleiotropy can impose significant constraints for adaptive traits, however they are not often quantified limiting understanding of how organisms may respond to their environment.

Our purpose for this study was to investigate the genetic determinants of drought related traits and their relationship to one another in an ecologically, economically and culturally important tree species. We used over 6 million SNPs and phenotyped 432 trees under common garden conditions for three drought-related traits to identify SNPs related to drought; we investigated trait heritability, genomic architecture, functional annotation, and gene interactions between multiple traits. We hypothesise that genomic architecture (epistasis and haplotype blocks) plays an important role in determining traits and that pleiotropy may constrain other drought-related physiological traits. We discover several significant genes, overabundance of significant loci in *cis*-regulatory regions, and many epistatic and pleiotropic interactions between significant

SNPs that may constrain drought traits in a foundational Australian tree species, *Corymbia calophylla*. These complex genomic architectures are likely to play important roles when managing natural and planted forests for drought under climate change.

Results and discussion

Species and functional traits

We quantified variation in three functional traits indicative of drought resistance (δ^{13}C, NDVI, SLA; Fig. 1b-d). Two of the three focal traits (δ^{13}C: $h^2=0.17^{**}$; NDVI: $h^2=0.15^{**}$; SLA $h^2=0.08$) exhibited small, but significant, narrow-sense heritabilities based on quantitative genetic models [29], indicating genetic determination by polygenic mechanisms. We controlled for random variation of the measured trait values using best linear unbiased predictions (BLUPs) and found that the three traits varied across families and populations (Fig. 1b-d). A linear model showed that traits were significantly differentiated among populations ($F_{11,419}=88.89$; $p<0.001$) and among families within populations ($F_{23,407}=53.75$; $p<0.001$).

Draft genome and genome-wide association studies (GWAS)

We assembled a high-quality de novo genome (350 Mb haploid size, 100 contigs, contig N50 = 3 Mb, 11 pseudochromosomes, NCBI accession #: GCA_014182845.1).

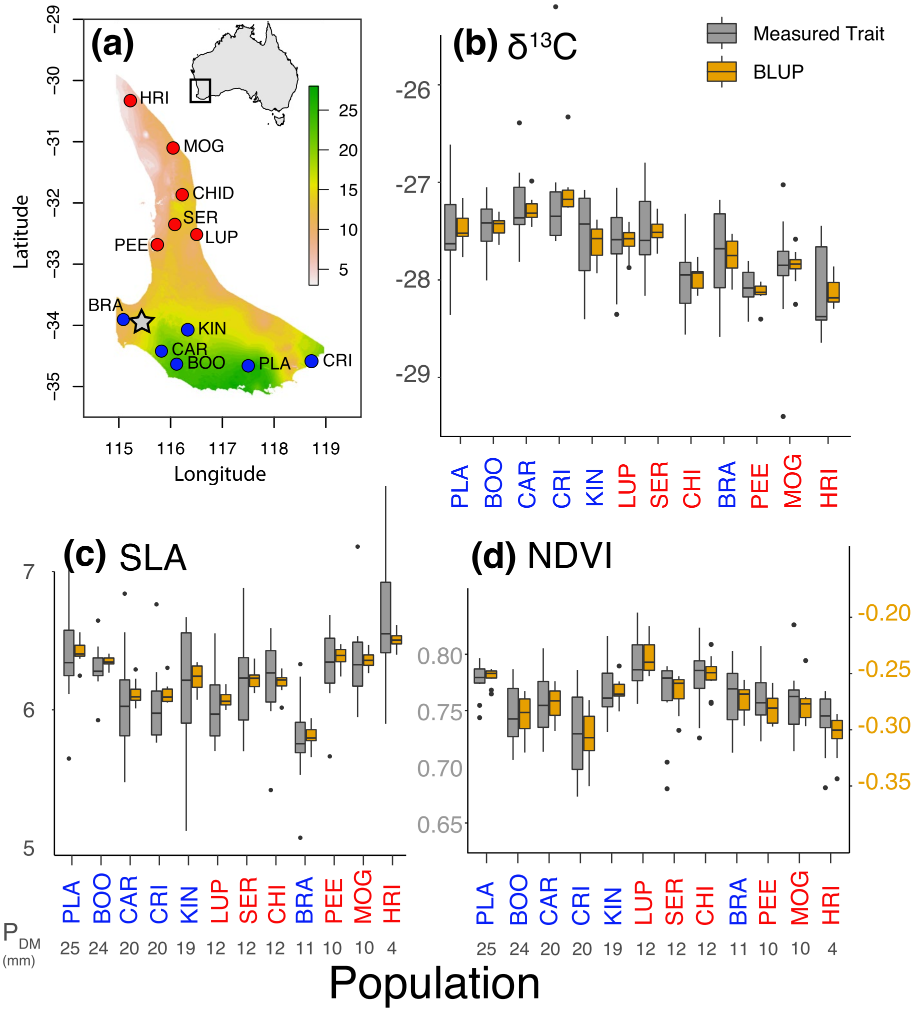

Fig. 1 Phenotypic traits were studied in a common garden of populations sampled from across the range of *Corymbia calophylla*, denoted here by the precipitation layer. **a** Location of populations sampled and the experimental site mapped with precipitation of the driest month (P_{DM}; mm; BIO14); **b**-**d** trait values (grey) with their best linear unbiased predictions (BLUPs; yellow) for (**b**) δ^{13}C, (**c**) SLA, and (**d**) NDVI. Population colours are coded red for northern populations and blue for southern populations and ordered from wettest (left) to driest (right). Star represents the location of the experimental site. Inset shows location of study area within Australia. NDVI was scaled (y2-axis; yellow) to meet assumptions of normality before estimating BLUPs

We then aligned individual short read sequences and identified 91 M pre-filtered single nucleotide polymorphisms (SNPs) (Figure S1; Table S1). First, low quality samples were removed if their missingness was below 0.5. Then variants were filtered on quality (Phred > 25), minimum read depth (6), missingness (max = 0.2), and allele frequency (minor allele frequency = 0.01), 6.49 million informative SNPs were discovered, averaging a SNP every 60 bases. Linkage disequilibrium (LD) decayed quickly, as median base pair distance to half-maximal r^2 values were 160 base pairs (Table S2). We also estimated the LD scores across each chromosome (Figure S2) and similar LD patterns for each chromosome. These mean LD estimate (160 bp) is greater than previous half-maximal estimates of LD decay in *Eucalyptus* species (92 and 113 bp) [39], confirming that there is a very high degree of population diversity and recombination in the system.

To discover SNP-trait associations, we performed genome wide association studies for the three functional traits. Weak but detectable population structure ($F_{ST} = 0.05$) was controlled for in the GWAS analysis using the first 10 axes of a multidimensional scaling plot, the first two axes show a distinction between the northern and southern populations (Figure S3). The resulting GWAS identified 279, 69 and 92 significant SNPs for δ^{13}C, SLA, and NDVI, respectively (Fig. 2a). Candidate SNPs were found on all chromosomes across the genome with several regions having a high density of candidate SNPs with peaks on chromosomes 3, 8, and 10 (Fig. 2a). Magnification of these peaks highlights many SNPs that occur in large haplotype blocks (150–350 kb) based on Haploview results, beyond the median LD decay, interspersed with non-significant SNPs, and different among all three traits (Fig. 2b). These patterns within gene-rich

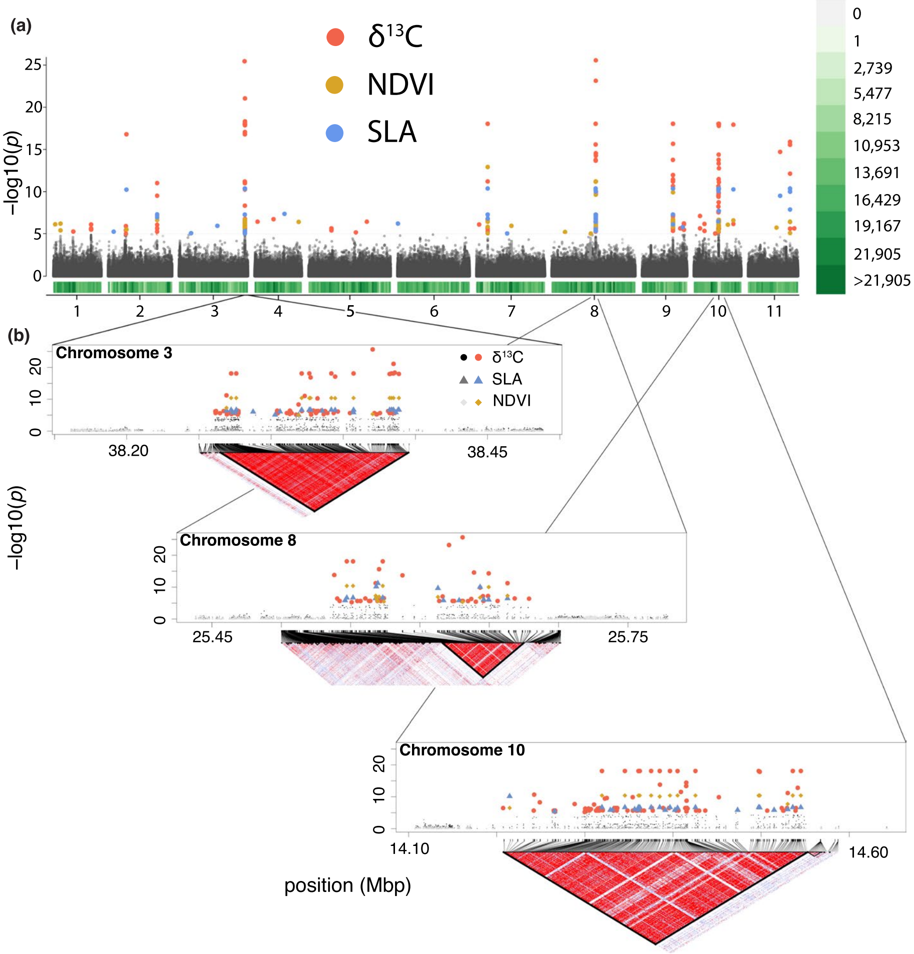

Fig. 2 Genome sampling and GWAS outputs for *Corymbia calophylla*. **a** Manhattan plots for three traits and SNP density. Points represent SNPs significantly associated with the trait (red = δ^{13}C; yellow = NDVI; blue = SLA; grey = not significant) at an FDR value < 0.00001. The density plot (below Manhattan plot) shows the number of SNPs in 1 million base pair segments across the genome with colour (white to green). **b** Magnified view of the significant peaks within the three 'hotspot' regions of adaptive variation. Underneath the magnified views are haploview plots that detect significant blocks, bounded by black lines. The linkage disequilibrium (r.^{2}) across the haploview plot is denoted from low (white) to high (red)

regions could be due to structural variants such as inversions [40] or large haplotype blocks, which have been found to be important in adaptation in other systems (sunflower ecotypes [41] and teosinte [42]). Structural variants are a significant source of variation contributing to adaptation [43], are non-randomly distributed throughout the genome [44], and can change gene expression patterns [45]. The patterns found, particularly the association of trait-associated SNPs with significant haploblocks, in our study could be due to structural variants, but long-read sequencing would have to be performed for confirmation. Large haplotype blocks likely explain the strong LD between candidate SNPs within chromosomes (Figure S4d-f) with significant LD for the haploblocks in chromosome 3 ($r^2=0.35$; $p=0.03$), chromosome 8 ($r^2=0.25$; $p=0.02$) and chromosome 10 ($r^2=0.29$; $p=0.04$) associated with δ^{13}C. However, these patterns do not explain any of the long-range LD across chromosomes (Figure S4a-c). The mean r^2 across all significant SNPs associated with δ^{13}C is 0.28 with a mean *p*-value of 0.04. This is likely due to rarity disequilibrium (i.e., genetic indistinguishability – giSNP [46]), which is a widespread phenomenon that arises when interchromosomal SNP pairs are in perfect LD due to the combinatorial limit on unique genotype patterns in finite sample sizes and may be contributing to the pattern. Even though we do not know the genomic mechanism of these haplotype blocks, we can be confident that the target genes within regions are indicative of the genetic architecture associated with quantitative traits due to high significance and extremely high LD between significant SNPs, and given the pattern persists after removal of the giSNPs.

To determine how much trait variation could be explained by all genomic SNPs, we estimated the SNP-based heritability to explain the total proportion of variance in phenotypes [47]. We found that SNP-based heritability for all three traits (δ^{13}C $h^2_{\text{SNP}}=0.55$ (SE 0.14); SLA $h^2_{\text{SNP}}=0.27$ (0.12); and NDVI $h^2_{\text{SNP}}=0.66$ (0.14)) was much greater than the heritabilities calculated through quantitative genetics methods ($h^2=0.11$ (0.08), 0.08 (0.08), and 0.15 (0.08) for δ^{13}C, SLA, and NDVI, respectively [20]), and genetic correlations (r_g) depended upon the traits. r_g was significant between SLA and NDVI ($p=0.008$) but not significant between δ^{13}C and the other two traits (SLA: $p=0.16$; NDVI: $p=0.55$). Even though both NDVI and SLA are weakly correlated wtih δ^{13}C at the trait level (SLA: $r^2=-0.2$; $p<0.001$; NDVI: $r^2=0.19$; $p<0.001$), both correlative patterns disappear at the genetic level. This paradox could indicate that correlational selection could be occurring among traits [48] or alternatively there is a complex system of gene reuse among traits that are difficult to detect [49]. Indeed, complex evolutionary patterns have been observed in eucalypts where the same gene is reused in diverging ways under the same selection environments [50]. There were also major differences between the three h^2_{SNP} estimates. Considering the large h^2_{SNP} for both δ^{13}C and NDVI, which includes all SNPs, we also identified several SNPs with large effect sizes (top SNP for δ^{13}C 0.25, SLA 0.34, NDVI 0.17; Table S3). Theoretical work indicates that in highly polygenic traits, alleles with very small effect sizes could be ephemeral because they are prone to swamping by gene flow as different genotypic combinations can provide optimum fitness [51]. While these ephemeral, swamping prone SNPs, could contribute to our traits, they would be impossible to differentiate between neutral alleles in a GWAS framework. On the other side of the effect-size spectrum, the top 10 candidate SNPs associated with δ^{13}C showed greater effects (18—25%) (Table S3; giSNPs were dropped from this model), compared to the greatest explanatory SNPs for NDVI and eight of the top 10 SNPs for SLA. These ten SNPs accounted for ~ 50% of the variation for δ^{13}C and SLA, and 34% of the variation in NDVI. The lower combined r^2 value for NDVI compared to the other two traits might be due to many factors, including more trait variation, more SNPs of small effect, and more epistatic interactions. The inclusion of epistatic interactions increased the variation explained for δ^{13}C and NDVI, but slightly lowered the variation explained for SLA, such that epistatic interactions improve phenotypic predictions for two of the three traits compared to individual SNP effects.

There is also a proportion of trait variation that was not explained. One explanation might be that these traits are highly polygenic with very small additive effects, and we were only able to identify the variants most strongly associated with the phenotype [52]. h^2_{SNP} should capture these undetected small effects, but this only explained a quarter to two thirds of the variation in our drought traits. Another explanation is that structural variants, which were not included in this dataset, could explain some of the missing heritability as they are known to be ubiquitous and have the potential to explain a large proportion of heritable genetic variation [55]. Even though we performed our experiment in a common garden to minimse the effect of the environment, phenotypic plasticity could still play a role in these trait differences through variation in gene methylation or control through regulatory elements among genotypes sourced from different environments [26, 53, 54]. Methylation could result in non-heritable differences, while variants in regulatory elements could result in differences in plastic responses that are heritable. We detected genomic variants in the regulatory region (details below), which explained a large proportion of the trait variation, despite complex interactions constraining adaptation, and provide insights into the adaptive role of genes and regulatory elements.

Complex gene and trait interactions (i.e., epistasis and pleiotropy) are known to play important roles in quantitative traits [56, 57]. We found evidence that some of the unexplained trait variation could be attributable to epistasis among significant SNPs identified in the GWAS. Gene interactions among the candidate SNPs were explicitly evaluated using CAPE (giSNPs are removed from this analysis), revealing significant epistatic interactions across the genome ($p < 0.05$; Fig. 3a), with strong interactions between chromosomes 3, 8, 9, and 10. Main effects between two SNP pairs are shown between chromosomes 3 & 7 (negative effect; blue arrows in 3a) and 9 & 11 (positive effect; yellow arrows in 3a) (Fig. 3b), and we also provide visualisation of an epistatic interaction when the main effect (variant + trait) is conditioned on a second variant (Fig. 3c), where the interaction between the two SNPs affects the trait in a negative way (Fig. 3c dashed line). We then assessed possible pleiotropic interactions between the three traits, i.e., the effect of one SNP on multiple traits. Pleiotropic interactions are shown in Fig. 3a in the concentric bands, where the same SNP is highlighted for more than one trait. We identified several cases within chromosomes 1, 3, 7, 8, 9, and 10 that were

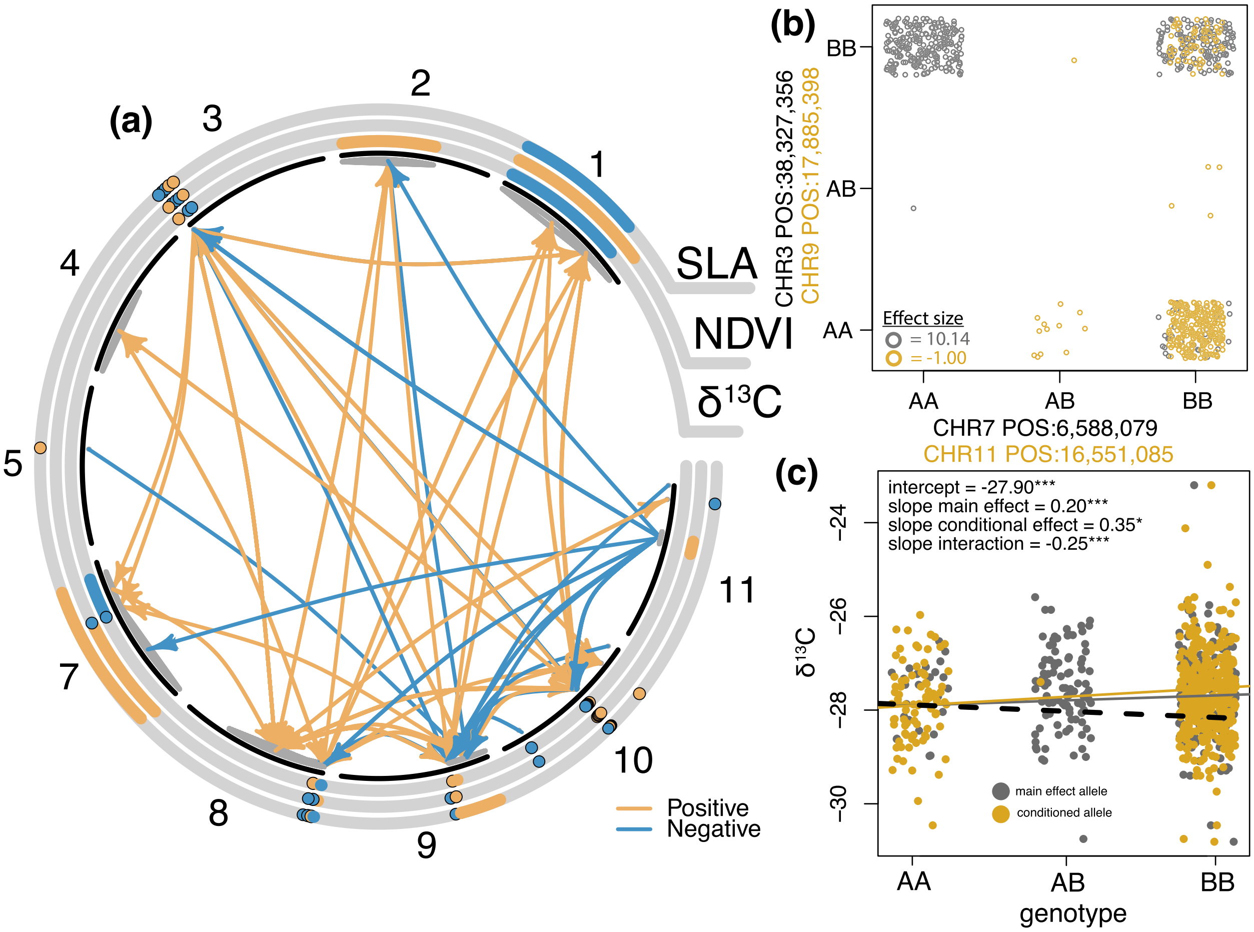

Fig. 3 Patterns of significant epistatic and pleiotropic interactions in *Corymbia calophylla*. **a** Epistatic interactions are shown with coloured arrows and pleiotropic effects between traits are shown in the circular bands. Chromosomes are in black; chromosome six is not present because no SNPs were significant in the analysis. The direction of influence is shown by colour, where orange indicates that the SNP affects a different SNP in a positive way and blue is indicative of a negative effect. Interactions between a SNP and multiple traits indicate pleiotropy, while the same colours are indicative of the same effects. Antagonistic pleiotropy is inferred if the colours are different among SNPs in the same chromosomal location. Points on chromosome 3, 8, and 10 have been manually separated due to severe overlapping to visualise the antagonistic effects. **b** Genotypes for negatively influenced epistatic interaction between two SNP variants (grey points in (**b**) & blue arrows in (**a**)) and positively influenced epistatic interactions between two SNP variants (orange points in (**b**) & orange arrows in (**a**)). **c** Visualisation of one significant epistatic interaction where the main effect of a SNP (grey in (**c**)) and trait is conditioned on a second SNP (yellow in (**c**)), black dashed line is the interaction effect between the two variants. *P*-value – *** < 0.001; * < 0.05

due to pleiotropic effects (Figs. 3a & S3). There was evidence of antagonistic interactions (when a gene affects traits in different directions; represented by different colours along the concentric trait-circles in Figs. 3a & S3). A network plot shows the interactions (both positive and antagonistic) between all 11 chromosomes (Figure S5). Pleiotropy was corroborated through annotation results, for example, there were 11 genes that were found to be significantly associated with all three traits (Table S4). Eight of these genes were expressed during growth and development processes, including in plant organs such as guard cell and leaf structure. Pleiotropy is known to play in integral role among correlated polygenic traits. In fact, a recent study on humans shows that 90% of trait-associated loci overlap with other traits and are mostly involved in the regulation of transcripts [58]. It is also known that pleiotropic loci maintain stronger genetic correlations compared to loci in LD [59]. This suggests that the interplay between pleiotropy, regulatory regions, linkage, and r_g is a critical component to tease apart the genetic mechanisms controlling drought tolerance in eucalypts. When assessing epistatic and pleiotropic interactions, it is difficult to determine how these control functional traits as there are tens of thousands of pairwise possibilities in our dataset contributing to the overall pattern of adaptation.

Annotation

In order to identify location (e.g., *cis* regulatory, genic) and effect (e.g., synonymous, nonsynonymous) of adaptation, snpEFF was used to annotate all SNPs. Annotations for all 6.49 million filtered SNPs reveal many moderate (nonsynonymous) and low (synonymous) effect alleles on protein function with a much smaller proportion of high-effect alleles (Table S5). Each chromosome had similar rates of nonsynonymous and synonymous SNPs (Table S5), except for chromosome 8 with a much higher rate of SNPs upstream and downstream of genes and more than double the number of high-effect SNPs than the other chromosomes (Table S5). We then functionally annotated the candidate SNPs within the three chromosomes with the highest significant peaks show interesting patterns (Chromosomes 3, 8, and 10; red in Fig. 2a). For example, of the 41 SNPs on chromosome 8 associated with $\delta^{13}C$, 37 are within gene regulatory regions (< 5,000 base pairs upstream of the gene), while the four remaining SNPs were nonsynonymous with a moderate effect and synonymous with low effects (Table 1). We should be cautious in extrapolating this further because of the non-independence between significant SNPs identified in the GWAS analysis within a haplotype block and further work should be performed to identify the causal SNP(s). However, the general overabundance of significant SNPs within regulatory regions compared to genic or intergenic regions is suggestive that regulatory regions play an important role. The regions of adaptive variation in chromosomes 3 and 10 are mostly intergenic with SNPs in gene *cis*-regulatory regions (within 5 kb of a gene on either the 5' (upstream) or 3' (downstream) end of the gene) and six candidate SNPs in promoter regions (within 500 bp upstream of a gene) for $\delta^{13}C$ on chromosome 10. This is notable because sequence variation in regulatory regions differentially impacts the function of nearby genes [60].

There were several significantly associated SNPs enriching genes with functions that provide support for the potential contribution to drought response (Table S6; results from orthofinder and eggNOG). For example, two genes were enriched for lignification and F-box protein (Eucgr.H02869 [61] and Eucgr.H02864 [62] respectively), which are known to support drought tolerance. In addition, the gene Eucgr.D00100 regulates the ethylene hormone, and is an ortholog to the Arabidopsis gene

Table 1 Annotation summary for the candidate SNPs on three chromosomes in *Corymbia calophylla*

Chromosome	Trait	SNPs	Intergenic	Up-reg (5 kb)	Prom (500 bp)	NS	S	Down-reg (5 kb)	Effect (H|M|L)
3	SLA	17	14	1	0	0	0	2	0|0|0
	NDVI	23	18	2	0	0	0	3	0|0|0
	$\delta^{13}C$	74	59	6	0	1	0	8	0|1|0
8	SLA	13	0	13	6	0	0	0	0|0|0
	NDVI	16	0	16	5	0	0	0	0|0|0
	$\delta^{13}C$	41	0	37	25	1	3	0	0|1|3
10	SLA	16	5	8	0	0	0	3	0|0|0
	NDVI	26	8	11	0	0	0	7	0|0|0
	$\delta^{13}C$	88	28	37	6	2	0	19	0|2|0

up-reg upstream regulatory region, *down-reg* downstream regulatory region, *prom* promoter region (within 500 bp of a gene), *NS* Nonsynonymous, *S* Synonymous, *H* High effect size (highly disruptive impact on protein function), *M* Moderate effect size (non-synonymous mutations, possible change in protein effectiveness), *L* Low effect size (synonymous mutations, non-coding or intergenic variant), *Intergenic* SNP not found within 5 kb of a gene

AT4G20880. Ethylene is known to mitigate the negative effects of water and temperature stresses [63]. The Eucgr. D00030 gene is an Ankyrin repeat family protein and has been known to confer tolerance to both drought and salinity in Arabidopsis and Soybean [64]. The possible roles these genes play in drought adaptation for *C. calophylla* will need to be quantitatively verified, but these discoveries provide promising ways in which this species have evolved drought resilience.

Cis-regulatory variants drive trait adaptation

The adaptive variation associated with traits, particularly for δ^{13}C, is largely driven by variants in *cis*-regulatory regions (noncoding DNA that regulates neighbouring genes; Table S7 – categorised significantly associated SNPs into four categories 500 bp, 5 kb, 10 kb and 50 kb), which are less constrained by pleiotropy than coding regions from an evolutionary perspective [65]; this mechanism appears to be important in *C. calophylla*. Indeed, recent studies suggest that *cis*-regulatory regions are critical for different types of adaptation [66–68]. Yet there is poor understanding how this variation influences population-level local adaptation, as noted by recent studies on evolution [69]. Here, we characterise variants associated with functional traits that are important for this species' adaptation to drought that are overrepresented by *cis*-regulatory regions. While we recognise that this finding needs to be confirmed in future research to disentangle non-independence issues within haplotype blocks, our data suggests that adaptation within *cis*-regulatory regions are more abundant than variants found within protein-coding genes and are more likely to shape the genomic architecture of these drought traits. Similarly, recent work has shown that regulatory variants are critical for drought in sunflowers [70]. We currently hypothesise that intraspecific drought-related phenotypes is mostly governed by changes within regulatory regions.

Conclusion

Considering the impact climate change is having on drought frequency and severity, understanding the molecular underpinnings of drought related traits provides an important step forward in determining the mechanisms controlling drought tolerance. We found heritable genetic variation associated with drought traits within several haplotype blocks across several chromosomes. This is particularly important when considering the abundance of epistatic and pleiotropic interactions, which likely constrain these traits ability to adapt. Furthermore, the majority of significant variants were detected in regulatory regions where they may influence the expression of many genes and traits. Despite the moderate levels of heritable variation determining drought related traits, the complex genomic architecture will complicate adaptive management strategies, i.e., by promoting one trait or gene, other traits or genes may be unexpectedly promoted or suppressed. Using the standing genomic variation in highly admixed natural populations may facilitate adaptation to climate change induced droughts.

Methods and materials

Study species

Corymbia calophylla is a foundation forest canopy species located in Western Australia (WA). It is considered a foundation species because it is critical for forest structure and ecological processes [71]. *Corymbia calophylla* is an important component of planted forests both for wood production and ecological restoration, provides critical habitat and resources to native animals, as well as having deep connections to the Aboriginal people. This species is an ideal candidate in which to study adaptation of functional traits because its distribution traverses strong environmental gradients over short distances, it has recently experienced mortality events attributed to climate change [37, 72], and evidence of adaptation to climate has been identified in physiological experiments and genome–environment investigations [20, 38, 73, 74].

Experimental site

This research was conducted in a plantation near Margaret River, WA Australia (Fig. 1 main text), located in the *C. calophylla*'s cool–wet region. Seed collection and trial design have been described in detail elsewhere [38]. Briefly, 18 populations represented by 165 families were established at the experimental site for a total of 3,960 individuals in six replicated blocks with two rows of buffer trees to minimise edge-effects. Seed collections for field trials were performed by Richard Mazanec (WA Department of Biodiversity, Conservation and attractions) and no voucher specimens were collected because the field sites are persistent. Families are defined here as individuals that have a known, common mother but unknown fathers (i.e., half-sibs) via mixed pollination within an intact forest. We focused on 12 populations representing contrasting climate combinations covering the full geographic distribution of *C. calophylla* (Fig. 1 main text). We sampled phenotypes and genotypes from a total of 432 trees, including 4 half-sibs from 10 families within 12 populations when available for a total of 120 families. Permissions for leaf material collection were provided by the land owners and Western Australia's Department of Biodiversity, Conservation, and Attractions.

Trait measurements

Traits were measured in March 2017 on *C. calophylla* trees that were 29 months old and 2–3 m tall. For each individual tree, we removed a north facing, mid-canopy side branch at its intersection with the main stem. The side branch was removed in the morning (between 8 a.m. and 12 noon), stored in a cool box, and measured in the afternoon (between 12 noon and 6 p.m.). For each side branch, we collected data for the three traits (among others not listed here): integrated water-use efficiency (δ^{13}C), specific leaf area (SLA), and normalized difference vegetation index (NDVI). All traits have shown close association to climate in past studies. High water-use efficiency (WUE) is the link between photosynthesis and evaporation [75] that translates to climatic tolerance under water limitation. Water-use efficiency is correlated with isotope discrimination (δ^{13}C, an isotopic signature measuring the ratio of 13C and 12C [76]) and relates to leaf gas exchange properties [77, 78]. To estimate δ^{13}C, the leaves were kept in an airtight box with silica gel until they could be dried in an oven at 70 °C for 48 h. δ^{13}C was measured from leaves dried using a benchtop freeze dryer (Alpha 1–4 LDplus Laboratory Freeze Dryer, Martin Christ). The leaves were grounded into a fine powder using a cyclotec mill (Foss Analytics) and sent for isotope analysis (ANU Isotope Laboratory) using a coupled EA-MS system (EA 1110 Carlo Erba; Micromass Isochrom).

Leaf-level normalized difference vegetation index (NDVI), which is generally used to measure chlorophyll content by quantifying leaf greenness, and is closely related to fraction of absorbed photosynthetically active radiation (FPAR) [79, 80]. While not technically a functional trait (NDVI), traits based on spectral properties of leaves can be indicative of photosynthetic activity and plant stress, and from hereon, we include this complex trait as a functional trait for ease of discussion. A field spectroradiometer (ASD standard-resolution FieldSpec4, Malvern Panalytical) was used to measure leaf reflectance in the visible and reflected infrared spectral regions with 2,151 narrow bands (10 nm full width at half maximum) and 1 nm spacing between band centers. Measurements were made for three leaves using a leaf-clip attachment with its own light source and calibrated to % reflectance using data collected from a Spectralon white reference panel. Means for all bands among the three leaves were calculated for each individual tree. Specific wavelengths were used to estimate the modified red-edge NDVI. The modified red-edge NDVI was calculated using the following equation [81]:

$$\mathrm{mND}_{705} = (\mathrm{R}_{750} - \mathrm{R}_{705})/(\mathrm{R}_{750} + \mathrm{R}_{705} - 2 \times \mathrm{R}_{445})$$

and was developed as an improvement to the standard NDVI to provide a more robust estimate of chlorophyll content [82] across a wide range of species and leaf structures [81]. Henceforth, this index will be referred to as "NDVI" in the text.

Specific leaf area (SLA) varies across global climate gradients [83], and high SLA values increase tree susceptibility to drought-induced mortality [84]. Specific leaf area (SLA) was measured on three fully matured leaves that were representative of the branch. After removing half of the petiole with a razor, the leaves were scanned into a computer using a Canon flatbed scanner (model # LiDE220) at 50 dpi. The leaves were then dried in an oven at 70 °C for 48 h and leaf mass was estimated on a digital scale with 1000th of a gram accuracy. SLA was calculated by dividing total leaf area by total leaf mass for all three leaves and averaged across the three leaves for a single SLA value for each individual tree.

BLUP estimation

Best linear unbiased predictions (BLUP) were estimated for each trait to account for variation attributed to the design matrix and to increase trait accuracy because it anticipates regression of progeny to the mean observed [85]. Analysis was performed using ASreml Version 4.1 [86, 87]. Univariate variances were estimated within the framework of the linear mixed model:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Z}\boldsymbol{u} + \mathbf{e}$$

Where $\mathbf{Y}$ is the column vector of individual phenotypic values of the response variable, $\mathbf{X}$ is the design matrix associating observations with fixed effects, $\boldsymbol{b}$ is the vector of fixed effects, $\mathbf{Z}$ is the design matrix associating observations with random effects, $\boldsymbol{u}$ is a vector of random effects and $\mathbf{e}$ is the vector of residual errors assumed to be identically and independently normally distributed with $E(e) = 0$.

Two sets of analysis were conducted, the first at the family level for the purpose of checking the data for homoscedasticity and determining if there was a need for transformation, and the second at the individual tree level for the purpose of estimating BLUPs.

Univariate family model

Elements in $\boldsymbol{b}$ included the intercept and provenance effects while elements in $\boldsymbol{u}$ included replicate, row within replicate, column within replicate, plot and family. Residual plots were examined for homoscedasticity and appropriate transformations identified as outlined previously [86, 87]. The trait NDVI was log transformed, whereas the δ^{13}C and SLA did not require transformation.

Univariate individual tree model

Elements in $\boldsymbol{b}$ and $\boldsymbol{u}$ were the same as for the univariate family model, with the exception that the family term was substituted with an individual tree, random additive effect. In this model, additive genetic covariance between relatives is modelled via the numerator relationship matrix. A one-tailed log likelihood ratio test with 0.5 degrees of freedom [87, 88] was used to test the significance of additive variance estimates for each trait.

Reference genome

DNA extraction and sequencing

We isolated high molecular weight DNA suitable for long-read re-sequencing by following a nuclei and magnetic bead-based extraction protocol [89]. Briefly, 30 g of fresh leaf material from an individual from the Australian Botanic Gardens in Canberra Australia was processed with 150 ml nuclei isolation buffer using a high-powered blender. The homogenate was filtered using a funnel of Miracloth. Next, 100% Triton X-100 was added to extract the nuclei from chloroplasts. The nuclei pellet was washed twice with a chilled nuclei buffer. Nuclei pellet lysis was performed with a lysis buffer containing 3% Sodium dodecyl sulphate (SDS) followed by incubating at 50°C. The DNA was cleaned of proteins by adding potassium acetate and pelleting. The supernatant was bound to Sera-MagTM SpeedBead magnetic carboxylate-modified particles (GE Healthcare). The beads were washed with 70% ethanol until clean. Size selection for fragments ≥ 30 kb was performed using a PippinHT (Sage Science, Beverly MA). MinION Mk1B was used to sequence the long-reads (Oxford Nanopore Technologies, ONT).

Nuclear genome assembly

Raw read libraries were filtered and trimmed in preparation of assembly with NanoPack [90] (NanoLyse version 1.1.0; NanoFilt version 2.6.0). First, ONT DNA control strand was removed. Next, 200 bp was trimmed from both 5' and 3' ends, removing sequencing adapters and low quality read ends. Finally, filtering removed all reads less than an average quality of 7 and less than 1 Kbp in length. Quality controlled read libraries were de novo assembled using the long read assembler Canu [91] (version 1.9; parameters: `corOutCoverage`=200 "`batOptions`=-dg 3 -db 3 -dr 1 -ca 500 -cp 50", `correctedErrorRate`=0.154, `corMaxEvidenceErate`=0.15, `–fast`). Following assembly, contaminant contigs were identified with blastn [92] (version 2.9.0+) using the NCBI nucleotide database [93] (versions BLASTDBv5). Identified contaminant contigs were removed with Blobtools [94] (version 1.1.1). Haplotigs, assembly artifacts, and plastid contigs were removed from assemblies with purge haplotigs [95] (version 1.1.0). Next, all assemblies were polished with the long read polisher Racon [96] (version 1.4.11) combined with minimap2 and the short read polisher Pilon [97] (version 1.23) combined with BWA-MEM [98]. Contigs of less than 1 Kbp were removed and manual curation of all remaining contigs was performed with MUMmer [99] (version 4.0.0beta2) to identify plastid DNA. Finally, our assemblies were scaffolded by RaGOO [100] using synteny information provided by the previously published *Eucalyptus grandis* genome [101]. Genome completeness was assessed with BUSCO [102] (version 3.0.2) and Lai [103] (version beta3.2). See Figure S1 for a summary of our assembly statistics.

Chloroplast assembly

Chloroplast reads were identified and subsequently extracted by aligning all reads with minimap2 [104] against a composite chloroplast genome made up of all published *Eucalyptus* chloroplast genomes. Chloroplast reads were identified within the alignment file with samtools v1.9 view [105] and extracted from all curated ONT reads using seqtk subseq (version: 1.3-r106; [106]). 1,000 chloroplast reads were randomly sampled using seqtk sample and assembled with Unicycler [107] (version 0.4.8) and polished using Pilon [97]. To confirm that the assembled contig was a chloroplast genome, an alignment dot plot was made of our chloroplast genome to the published *Eucalyptus* chloroplast genomes, using MUMmer [99].

Repeat and gene annotation

Prior to gene annotation repetitive regions in *C. calophylla's* genome were identified and soft masked with RepeatMasker [108] (version 4.1.1) using de novo repeat libraries created with EDTA [109] (version 1.9.6). Protein-coding genes and transcripts were predicted by BRAKER2 [110] (version 2.1.5; parameters: `epmode`), using 306,675 proteins sequences from Myrtaceae (Taxonomy ID: 3931) and 371,118 proteins sequences from *Arabidopsis thaliana* (Taxonomy ID: 3702) obtained from the National Center for Biotechnology Information [93] as homology evidence.

Library preparation & variant calling

DNA extraction for whole genome sequencing was performed by the Australian Genomic Research Facility (AGRF, Adelaide, SA Australia) using a modified CTAB method [111]. We generated short-read whole-genome shotgun DNA sequencing libraries using a low-cost transposase-based protocol [89]. Briefly, we quantified DNA concentrations using a fluorometric Quant-iT™ high sensitivity dsDNA assay kit (Molecular Probes™ Q33120). To normalise concentrations among samples, we diluted DNA to 2 ng/µl, quantified again and then diluted to 0.8 ng/µl. To form sequencing libraries, we combined 3 µl of each sample (approx 2.24 ng) with a small quantity of a Nextera™ tagment DNA enzyme (Illumina catalogue #15,027,865). To decrease costs, we performed this tagmentation reaction at 1/5th volume and 1/5th concentration of the manufacturer's protocol, i.e., 1/25th reactions. We amplified the libraries and added custom index sequences during 13 cycles of PCR. We purified and size-selected libraries using two SPRI-bead based cleanups and electrophoresis-based final size selection for insert sizes between 200 and 500 bp. We sequenced these libraries on a single S4 flow-cell on an Illumina NovoSeq 6000 instrument at Genomics West Australia/Telethon Kids Institute, Perth, West Australia.

Sequencing yielded between 3 and 10 Gbp per sample (~ 10-30X coverage), pooled across all sequencing runs (see Fig. 1 in main text). We discovered genetic variation among samples following a previous approach [39]. Briefly, we filtered, trimmed, and merged pairs of raw sequencing data using AdapterRemoval [112], then aligned reads to reference genomes using BWA-MEM version 0.7.15 [98, 113]. We detected short genomic variants using bcftools mpileup, normalised variants with bcftools norm, and performed initial variant filtering with bcftools filter [114]. Reads were aligned against our custom *Corymbia calophylla* reference genome. During initial variant filtering, we discarded variants with quality < 25, fewer than five reads in total across all alleles in all samples and fewer than three reads supporting the alternate allele across all samples. Resulting in 91 million pre-filtered single nucleotide variants, a variant every ~ 4 bp, which is normal among eucalypt species [39].

Filtering

After variants were called using the above pipeline, additional filtering was performed in PLINK 2.0 [115] with the following thresholds. Minimum read-depth was set to six. We extracted biallelic variants only, to ensure all variants were biallelic and minimise complex signals. The minimum basepair distance between variants was set to 10. Minor allele frequency (MAF) was set to 0.01, to have sufficient power for GWAS detection. Missing data threshold was set to 0.5 but the average missing data in the data set was 0.2. Resulting in a dataset with 6.5 million SNPs across all 11 chromosomes.

Linkage disequilibrium

Linkage disequilibrium (LD) was measured using median base pair distance to half-maximal r^2 values using boringLD v0.3.0 (https://github.com/kdmurray91/ boringld). We set the window size to 30 kbp with a 15 kbp overlap. We fitted analytical models of the decay of r^2 as a function of inter-SNP base pair distance using formulae derived by Hill and Weir [116] and then calculated base pair distance to half-maximal r^2 for each window. We summarized per-window estimates of half-maximal r^2 across all genome windows for a global r^2 estimate. To test if LD was a function of the number of SNPs within each window, we used a linear model within each chromosome (Figure S2). The linear fit was significant for all chromosomes but the r^2 values were low, this pattern was driven by the windows with very few SNPs. We also use LDSC to obtain LD scores which are the cumulative sum of r^2 values across SNPs within 30 kb windows [117] and plotted these scores for each chromosome using ggplot2 [118] and R [119].

Associations

Genome wide association studies (GWAS) were performed in Plink2 for each of the three functional traits. We used the individual BLUP estimates as the functional trait inputs, as this accounted for experimental site effects. We used the first 10 axes from an MDS as a covariate for population structure (first two axes are plotted and shown in Figure S3). We used the general linear model (glm) function to calculate *p*-values. We note that the power of GWAS analyses increases with higher genetic variation [120], but *Corymbia calophylla* is known to have extremely high diversity across its range with high connectivity [74, 121], making this species an exceptional study organism for this type of analysis. We also tested the associations of the GWAS using GEMMA with kinship matrix as a random covariable and the outputs resulted in nearly identical *p*-value distributions to the Plink2 analysis (Figure S6). We imported the Plink2 results to R [119] and adjusted the p-values for multiple comparisons using the Benjamini–Hochberg method (BH). Then used the *CMplot* command from the *CMplot* package to visualise the *p*-value distribution in a manhattan plot format. To ensure that the significant associations identified using the GWAS approach were not random, we used 100 permutations across the phenotype data in two ways and ran the Plink2 GLM analysis for both permutation variations. The first was completely random using the `sample` function in R without replacement, and the second was keeping the family structure of phenotypes and resampling among families.

The qq plots suggest that the *p*-values were inflated regardless of the covariable used (population or family structure; Figure S6). Both GWAS methods identified very similar groupings of SNPs (51–65% total similarity and top 50 SNPs were 88% similar for δ^{13}C). The differences between the distribution of BH adjusted *p*-values were significantly different between the real and permuted datasets (t.test: $T_{1,1.3\ m}$ = -2648.8, versus Random $p < 0.001$; $T_{1,1.3\ m}$ = -244.4, versus Family $p < 0.001$), suggesting that the adaptive variants were not due to chance.

We wanted to determine if the associations between SNP and trait were associated with local genomic structure, as described in Li and Ralph [122]. Therefore, we used the package `lostruct` in R to investigate the structure within 1000 SNP windows (this is equivalent to approximately 10 kbp) within each chromosome, creating between 400 and 600 windows per chromosome. Then we compared the first two axes within chromosomes to the location of adaptive genomic regions for three major areas of association on chromosome 3, 8, and 10. We found that anomalous local population structure among 1000 SNP windows was not localised near regions that were significantly associated with phenotypes (Figure S7). Significant haplotype blocks within these three major regions were identified using Haploview [123] with 500 max kb, 0.05 minor allele frequency threshold, and block significance was determined using the default option of 95% confidence intervals [124].

SNP heritability

To calculate SNP based heritability, we used the GEMMA model to describe the proportion of variance in phenotypes explained (PVE). GEMMA fits a univariate linear mixed model for marker association tests with a single phenotype, and for estimating the PVE by all variants [125]. We acknowledge that this work is performed in one common garden and that shared environments are known to inflate heritability [126]. However, upward bias due to shared environments would be consistent across populations [127] such that the heritability relationships are comparable among populations within the study species. This difference between actual and inflated heritability indicates that not all variation identified is adaptive, further quantitative experiments would need to be performed to confirm these results. We also determined genetic correlation (r_g) between the three traits using LDSC [128], following the author's recommendations.

Epistasis & pleiotropy

We attempted to uncover some of the complex epistatic and pleiotropic relationships between variants and traits by use of the combined analysis of pleiotropy and epistasis (CAPE) package in R [129], which implements an analytical method described in Carter et al. [130] to explicitly test for these complex interactions. This method was designed for datasets that include populations with mixed genetic variation, and is therefore appropriate for our study design. CAPE calculates both the main effects, which are the effect of a SNP from the set of all pairwise regressions that included that SNP, and the directional influences of that SNP that interact epistatically. We used the following parameters for the CAPE analysis (parameter file available online): `traits_scaled`=true, `pval_correction`=fdr, `alpha`=0.5, `peak_density`=0.8, `tolerance`=10, `num_alleles_in_pairscan`=300, `maxpair_cor`=0.5, `pairscan_null_size`=1000. We used a high `peak_density` because of the quick LD decay, as suggested in the CAPE documentation. We also used a `num_alleles_in_pairscan` of 300 to limit the number of SNP pair analyses, this results in a different outcome for each run because we do not test all 104,329 pair possibilities. To be clear, individual SNP pair outcomes will not change, it is whether or not the individual SNP pair is randomly included in the output. Even so, the result shown here is a representative subset of these interactive effects. Both the inputs and outputs for our specific CAPE analysis are provided online, so the user can recreate our figures but also explore other individual runs and create new figures.

Functional annotations

The program snpEFF [131] was used to identify the location of significantly associated SNPs using the *Corymbia calophylla* genome (NCBI txid34324; assembly ASM1418284v1). Variants found within genes were recorded as synonymous or nonsynonymous, in addition variants in regulatory regions found within 5,000 base pairs of genes were recorded as being upstream or downstream, along with the number of base pairs between the gene and SNP. We recorded putative impact of the SNP on gene function and generally moderate effects are from nonsynonymous SNPs (changes in amino acids; 'M' in Table 1), low effects are from synonymous SNPs (no changes to amino acids; 'L' in Table 1), and high effects are from frame shifts or changes to start/stop codons (loss off function; 'H' in Table 1). We also specified which variants are in promoter regions, defined here as being within 500 bp upstream of the gene. Then, orthofinder was used to identify homologs between *C. calophylla* and *E. grandis* genomes [132], and assign putative functions to predicted genes identified as significant. We also provided results from a scale mapper to identify possible orthologs across KEG, COG, and eggNOG databases using eggnog-mapper [133] (version: 5.0) using the sequence aligner diamond [134] (version: 2.0.15).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10531-8.

Additional file 1: Tables S1 – S5, S7. Figs S1 – S6.

Additional file 2: Table S6.

Acknowledgements
We thank Anthea Challis, Katinka Ruthroff, Giles Hardy, Villa Weber, Sarah Sapsford for assistance with field data collections. We thank Jason Bragg and Peter Harrison for insightful discussions about trait predictions.

Authors' contributions
CA and PR developed the project; CA, PR, MB, DT, and RM collected field data; laboratory work was supervised by AJ; reference genome was developed by SF; bioinformatics and analyses were performed by CA, KM, and RM; CA wrote the first draft of the manuscript, while CA, KM, RM, SF, JB, DT, MB, JOB, and PR edited the manuscript.

Funding
Australian Research Council Linkage Grant (LP150100936).

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the European Variation Archive (EVA) repository, project code – PRJEB67343 (https://www.ebi.ac.uk/eva/?eva-study=PRJEB67343).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 September 2023 Accepted: 14 June 2024
Published online: 27 June 2024

References

1. Konapala G, Mishra AK, Wada Y, Mann ME. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. Nat Commun. 2020;11:3044.
2. Harris RMB, Beaumont LJ, Vance TR, Tozer CR, Remenyi TA, Perkins-Kirkpatrick SE, et al. Biological responses to the press and pulse of climate trends and extreme events. Nat Clim Change. 2018;8:579–87.
3. Cuervo-Alarcon L, Arend M, Müller M, Sperisen C, Finkeldey R, Krutovsky KV. A candidate gene association analysis identifies SNPs potentially involved in drought tolerance in European beech (Fagus sylvatica L.). Sci Rep. 2021;11:2386.
4. Moran E, Lauder J, Musser C, Stathos A, Shu M. The genetics of drought tolerance in conifers. N Phytol. 2017;216:1034–48.
5. Songsomboon K, Brenton Z, Heuser J, Kresovich S, Shakoor N, Mockler T, et al. Genomic patterns of structural variation among diverse genotypes of Sorghum bicolor and a potential role for deletions in local adaptation. G3 GenesGenomesGenet. 2021;11:jkab154.
6. Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, et al. Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (Arachishypogaea L.). Theor Appl Genet. 2011;122:1119–32.
7. Varshney RK, Thudi M, Nayak SN, Gaur PM, Kashiwagi J, Krishnamurthy L, et al. Genetic dissection of drought tolerance in chickpea (Cicer arietinum L.). Theor Appl Genet. 2014;127:445–62.
8. Hoffmann AA, Hercus MJ. Environmental stress as an evolutionary force. Bioscience. 2000;50:217–26.
9. Hamann E, Weis AE, Franks SJ. Two decades of evolutionary changes in Brassica rapa in response to fluctuations in precipitation and severe drought. Evolution. 2018;72:2682–96.
10. Farquhar GD, Ehleringer JR, Hubick KT. Carbon isotope discrimination and photosynthesis. Annu Rev Plant Physiol Plant Mol Biol. 1989;40:503–37.
11. Condon AG, Richards RA, Rebetzke GJ, Farquhar GD. Improving intrinsic water-use efficiency and crop yield. Crop Sci. 2002;42:122–31.
12. Hubick K, Farquhar G, Shorter R. Correlation between water-use efficiency and carbon isotope discrimination in diverse peanut (Arachis) germplasm. Funct Plant Biol. 1986;13:803–16.
13. Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, et al. Genome-wide association study (GWAS) of carbon isotope ratio (δ13C) in diverse soybean [Glycine max (L.) Merr.] genotypes. Theor Appl Genet. 2015;128:73–91.
14. Torre ARDL, Sekhwal MK, Puiu D, Salzberg SL, Scott AD, Allen B, et al. Genome-wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. Plant J. 2021. https://doi.org/10.1111/tpj.15592.
15. de Miguel M, Rodríguez-Quilón I, Heuertz M, Hurel A, Grivet D, Jaramillo-Correa JP, et al. Polygenic adaptation and negative selection across traits, years and environments in a long-lived plant species (Pinus pinaster Ait., Pinaceae). Mol Ecol. 2022;31:2089–105.
16. Poorter H, Evans JR. Photosynthetic nitrogen-use efficiency of species that differ inherently in specific leaf area. Oecologia. 1998;116:26–37.
17. Wellstein C, Poschlod P, Gohlke A, Chelli S, Campetella G, Rosbakh S, et al. Effects of extreme drought on specific leaf area of grassland species: a meta-analysis of experimental studies in temperate and sub-Mediterranean systems. Global Change Biol. 2017;23:2473–81.
18. Chhetri HB, Macaya-Sanz D, Kainer D, Biswal AK, Evans LM, Chen J, et al. Multitrait genome-wide association analysis of Populus trichocarpa identifies key polymorphisms controlling morphological and physiological traits. New Phytol. 2019;223:293–309.
19. Shipley B. Plasticity in relative growth rate and its components following a change in irradiance. Plant Cell Environ. 2000;23:1207–16.
20. Ahrens CW, Andrew ME, Mazanec RA, Ruthrof KX, Challis A, Hardy G, et al. Plant functional traits differ in adaptability and are predicted to be differentially affected by climate change. Ecol Evol. 2020;10:232–48.
21. Karnieli A, Agam N, Pinker RT, Anderson M, Imhoff ML, Gutman GG, et al. Use of NDVI and land surface temperature for drought assessment: merits and limitations. J Climate. 2010;23:618–33.
22. Wang J, Li X, Guo T, Dzievit MJ, Yu X, Liu P, et al. Genetic dissection of seasonal vegetation index dynamics in maize through aerial based high-throughput phenotyping. Plant Genome. 2021;14:e20155.
23. Stinchcombe JR, Kelley JL, Conner JK. How to measure natural selection. Methods Ecol Evol. 2017;8:660–2.
24. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. Plos Genet. 2008;4:e1000008.
25. Marjoram P, Zubair A, Nuzhdin SV. Post-GWAS: where next? more samples, more SNPs or more biology?. Heredity. 2014;112:79–88.
26. Ahrens CW, Rymer PD, Tissue DT. Intra-specific trait variation remains hidden in the environment. New Phytol. 2021;229:1183–5.
27. Falconer D, Mackay T. Introduction to quantitative traits. 4th ed. London: Longman Group Ltd.; 1996.
28. Rouzic AL, Carlborg Ö. Evolutionary potential of hidden genetic variation. Trends Ecol Evol. 2008;23:33–7.
29. Challa S, Neelapu NRR. Biochemical, physiological and molecular avenues for combating abiotic stress tolerance in plants. Plant Gene. 2018;11:135–50.
30. Gupta PK, Kulwal PL, Jaiswal V. Chapter Two Association mapping in plants in the post-GWAS genomics era. Adv Genet. 2019;104:75–154.
31. Cortes LT, Zhang Z, Yu J. Status and prospects of genome-wide association studies in plants. Plant Genome. 2021;14:e20077.
32. Tyler AL, Emerson J, Kassaby BE, Wells AE, Philip VM, Carter GW. Epistasis. Methods Protoc Methods Mol Biol. 2021;2212:55–67.
33. Klei L, Luca D, Devlin B, Roeder K. Pleiotropy and principal components of heritability combine to increase power for association analysis. Genet Epidemiol. 2008;32:9–19.
34. Forsberg SKG, Bloom JS, Sadhu MJ, Kruglyak L, Carlborg Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat Genet. 2017;49:497–503.
35. Chen P, Zhang J. Antagonistic pleiotropy conceals molecular adaptations in changing environments. Nat Ecol Evol. 2020;4:461–9.
36. Rennison DJ, Peichel CL. Pleiotropy facilitates parallel adaptation in sticklebacks. Mol Ecol. 2022. https://doi.org/10.1111/mec.16335.

37. Matusick G, Ruthrof KX, Brouwers NC, Dell B, Hardy GStJ. Sudden forest canopy collapse corresponding with extreme drought and heat in a mediterranean-type eucalypt forest in southwestern Australia. Eur J Forest Res. 2013;132:497–510.
38. Ahrens CW, Mazanec RA, Paap T, Ruthrof KX, Challis A, Hardy G, et al. Adaptive variation for growth and resistance to a novel pathogen along climatic gradients in a foundation tree. Evol Appl. 2019;12:1178–90.
39. Murray KD, Janes JK, Jones A, Bothwell HM, Andrew RL, Borevitz JO. Landscape drivers of genomic diversity and divergence in woodland Eucalyptus. Mol Ecol. 2019;28:5232–47.
40. Wellenreuther M, Mérot C, Berdan E, Bernatchez L. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Mol Ecol. 2019;28:1203–9.
41. Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature. 2020;584:602–7.
42. Fang Z, Pyhäjärvi T, Weber AL, Dawe RK, Glaubitz JC, González de JJS, et al. Megabase-scale inversion polymorphism in the wild ancestor of Maize. Genetics. 2012;191:883–94.
43. Fuller ZL, Koury SA, Phadnis N, Schaeffer SW. How chromosomal rearrangements shape adaptation and speciation: case studies in Drosophila pseudoobscura and its sibling species Drosophila persimilis. Mol Ecol. 2019;28:1283–301.
44. Catanach A, Crowhurst R, Deng C, David C, Bernatchez L, Wellenreuther M. The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost Chrysophrys auratus. Mol Ecol. 2019;28:1210–23.
45. Hämälä T, Gorton AJ, Moeller DA, Tiffin P. Pleiotropy facilitates local adaptation to distant optima in common ragweed (Ambrosia artemisiifolia). Plos Genet. 2020;16:e1008707.
46. Skelly DA, Magwene PM, Stone EA. Sporadic, global linkage disequilibrium between unlinked segregating sites. Genetics. 2016;202:427–37.
47. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
48. Svensson EI, Arnold SJ, Bürger R, Csilléry K, Draghi J, Henshaw JM, et al. Correlational selection in the age of genomics. Nat Ecol Evol. 2021;5:562–73.
49. Westram AM, Galindo J, Rosenblad MA, Grahame JW, Panova M, Butlin RK. Do the same genes underlie parallel phenotypic divergence in different Littorina saxatilis populations?. Mol Ecol. 2014;23:4603–16.
50. Ahrens CW, Watson-Lazowski A, Huang G, Tissue DT, Rymer PD. The roles of divergent and parallel molecular evolution contributing to thermal adaptive strategies in trees. Plant Cell Environ. 2022. https://doi.org/10.1111/pce.14449.
51. Yeaman S. Local adaptation by Alleles of small effect. Am Nat. 2015;186:S74-89.
52. Láruson ÁJ, Yeaman S, Lotterhos KE. The importance of genetic redundancy in evolution. Trends Ecol Evol. 2020;35:809–22.
53. Corcuera L, Gil-Pelegrin E, Notivol E. Phenotypic plasticity in Pinus pinaster δ13C: environment modulates genetic variation. Ann Forest Sci. 2010;67:812–812.
54. Aubin-Horth N, Renn SCP. Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. Mol Ecol. 2009;18:3763–80.
55. Chakraborty M, Emerson JJ, Macdonald SJ, Long AD. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. Nat Commun. 2019;10:4872.
56. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat Rev Genet. 2014;15:22–33.
57. Yan W, Wang B, Chan E, Mitchell-Olds T. Genetic architecture and adaptation of flowering time among environments. N Phytol. 2021;230:1214–27.
58. Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019;51:1339–48.
59. Chebib J, Guillaume F. Pleiotropy or linkage? their relative contributions to the genetic correlation of quantitative traits and detection by multitrait GWA studies. Genetics. 2021;219:iyab159.
60. Riethoven J-JM. Computational biology of transcription factor binding. Methods Mol Biol. 2010;674:33–42.
61. Zhou Y, Zhang Y, Wang X, Han X, An Y, Lin S, et al. Root-specific NF-Y family transcription factor, PdNF-YB21, positively regulates root growth and drought resistance by abscisic acid-mediated indoylacetic acid transport in Populus. New Phytol. 2020;227:407–26.
62. Abd-Hamid N-A, Ahmad-Fauzi M-I, Zainal Z, Ismail I. Diverse and dynamic roles of F-box proteins in plant biology. Planta. 2020;251:68.
63. Chen H, Bullock DA, Alonso JM, Stepanova AN. To fight or to grow: the balancing role of ethylene in plant abiotic stress responses. Plants. 2021;11:33.
64. Zhao J-Y, Lu Z-W, Sun Y, Fang Z-W, Chen J, Zhou Y-B, et al. The ankyrin-repeat gene GmANK114 confers drought and salt tolerance in arabidopsis and soybean. Front Plant Sci. 2020;11:584167.
65. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* Enhancer. Science. 2010;327:302–5.
66. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007;17:1520–8.
67. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2012;13:59–69.
68. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015;16:197–212.
69. Lewis JJ, van der Burg KRL, Mazo-Vargas A, Reed RD. ChIP-seq-annotated heliconius erato genome highlights patterns of cis-regulatory evolution in Lepidoptera. Cell Rep. 2016;16:2855–63.
70. Todesco M, Bercovich N, Kim A, Imerovski I, Owens GL, Ruiz ÓD, et al. Genetic basis and dual adaptive role of floral pigmentation in sunflowers. Elife. 2022;11:e72072.
71. Ellison AM, Bank MS, Clinton BD, Colburn EA, Elliott K, Ford CR, et al. Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Front Ecol Environ. 2005;3:479–86.
72. Ruthrof KX, Matusick G, Hardy GEStJ. Early differential responses of co-dominant canopy species to sudden and severe drought in a mediterranean-climate type forest. Forests. 2015;6:2082–91.
73. Aspinwall MJ, Vårhammar A, Blackman CJ, Tjoelker MG, Ahrens C, Byrne M, et al. Adaptation and acclimation both influence photosynthetic and respiratory temperature responses in Corymbia calophylla. Tree Physiol. 2017;37:1095–112.
74. Ahrens CW, Byrne M, Rymer PD. Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. Mol Ecol. 2019;28:2502–16.
75. Yang Y, Guan H, Batelaan O, McVicar TR, Long D, Piao S, et al. Contrasting responses of water use efficiency to drought across global terrestrial ecosystems. Sci Rep-uk. 2016;6:23284.
76. Farquhar G, Richards R. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. Funct Plant Biol. 1984;11:539.
77. Diefendorf AF, Mueller KE, Wing ScottL, Koch PL, Freeman KH. Global patterns in leaf 13C discrimination and implications for studies of past and future climate. Proc National Acad Sci. 2010;107:5738–43.
78. Cernusak LA, Ubierna N, Winter K, Holtum JAM, Marshall JD, Farquhar GD. Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants. New Phytol. 2013;200:950–65.
79. Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens Environ. 2002;83:214–31.
80. Peng Y, Gitelson AA. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. Remote Sens Environ. 2012;117:440–8.
81. Sims DA, Gamon JA. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens Environ. 2002;81:337–54.
82. Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ. 1979;8:127–50.
83. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, et al. The worldwide leaf economics spectrum. Nature. 2004;428:821–7.
84. Greenwood S, Ruiz-Benito P, Martínez-Vilalta J, Lloret F, Kitzberger T, Allen CD, et al. Tree mortality across biomes is promoted by drought intensity, lower wood density and higher specific leaf area. Ecol Lett. 2017;20:539–53.

85. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica. 2008;161:209–28.
86. Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R, Butler D, et al. ASReml user guide release 4.1 structural specification. VSN International Ltd, 5. 2014
87. Gilmour A, Dutkowski G. Pedigree options in ASReml. Unpublished manuscript. 2004. https://www.animalgenome.org/bioinfo/resources/manuals/ASReml3/pedigree.pdf.
88. Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc. 1987;82:605–10.
89. Jones A, Torkel C, Stanley D, Nasim J, Borevitz J, Schwessinger B. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. PLoS ONE. 2021;16:e0253830.
90. Coster WD, D'Hert S, Schultz DT, Cruts M, Broeckhoven CV. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9.
91. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36.
92. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
93. Coordinators NR, Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2017;46:D8-13.
94. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000research. 2017;6:1287.
95. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460.
96. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.
97. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9:e112963.
98. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Arxiv. 2013: 1303.3997.
99. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinform. 2003;10.3:1–18.
100. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20:224.
101. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of Eucalyptus grandis. Nature. 2014;510:356–62.
102. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
103. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. 2018;46:e126.
104. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
105. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10:giab008.
106. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE. 2016;11:e0163962.
107. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. Plos Comput Biol. 2017;13:e1005595.
108. Tarailo-Graovac M, Chen N. Using repeatmasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 2009;25:410.1-4.10.14.
109. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275.
110. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. Nar Genom Bioinform. 2021;3:lqaa108.
111. Healey A, Furtado A, Cooper T, Henry RJ. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods. 2014;10:21.
112. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016;9:88.
113. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
114. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.
115. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:1–16.
116. Hill WG, Weir BS. Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol. 1988;33:54–78.
117. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.
118. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer; 2009. https://doi.org/10.1007/978-0-387-98141-3.
119. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
120. Hamazaki K, Kajiya-Kanegae H, Yamasaki M, Ebana K, Yabe S, Nakagawa H, et al. Choosing the optimal population for a genome-wide association study: A simulation of whole-genome sequences from rice. Plant Genome. 2020;13:e20005.
121. Sampson J, Tapper S, Coates D, Hankinson M, Mcarthur S, Byrne M. Persistence with episodic range expansion from the early Pleistocene: the distribution of genetic variation in the forest tree Corymbia calophylla (Myrtaceae) in south-western Australia. Biol J Linn Soc. 2018;123:545–60.
122. Li H, Ralph PL. Local PCA shows how the effect of population structure differs along the genome. Genetics. 2018;211:genetics.301747.2018.
123. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.
124. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225–9.
125. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.
126. Young AI, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, Sulem P, et al. Relatedness disequilibrium regression estimates heritability without environmental bias. Nat Genet. 2018;50:1304–10.
127. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. Plos Genet. 2013;9:e1003520.
128. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
129. Tyler AL, Lu W, Hendrick JJ, Philip VM, Carter GW. CAPE: An R Package for Combined analysis of pleiotropy and epistasis. Plos Comput Biol. 2013;9:e1003270.
130. Carter GW, Hays M, Sherman A, Galitski T. Use of pleiotropy to model genetic interactions in a population. Plos Genet. 2012;8:e1003010.
131. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly. 2012;6:80–92.
132. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
133. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:msab293.
134. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.

Publisher's Note