Research Article

Environmental drivers and genomic architecture of trait differentiation in fire-adapted
Banksia attenuata ecotypes

Running title: Trait differentiation in fire-adapted Banksia attenuata

Tianhua He1*, Byron B. Lamont1, Neal J. Enright2, Haylee M. D’Agui1, William Stock3
1. School of Molecular and Life Sciences, Curtin University, GPO Box U1987, Perth WA 6845, Australia
2. School of Veterinary and Life Sciences, Murdoch University, Perth WA 6150, Australia
3. Centre for Ecosystem Management, Edith Cowan University, Joondalup WA, 6027, Australia

*Correspondence: Tianhua He (Tianhua.He@curtin.edu.au)

Edited by: Randy Todd Clark, DuPont Pioneer, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/jipb.12697]

This article is protected by copyright. All rights reserved.
Received: February 9, 2018; Accepted: July 10, 2018
Abstract

Trait divergence between populations is considered an adaptive response to different environments, but to what extent this response is accompanied by genetic differentiation is less clear since it may be phenotypic plasticity. In this study, we analysed phenotypic variation between two *Banksia attenuata* growth forms, lignotuberous (shrub) and epicormic resprouting (tree), in fire-prone environments to identify the environmental factors that have driven this phenotypic divergence. We linked genotype with phenotype and traced candidate genes using differential gene expression analysis. Fire intervals determined the phenotypic divergence between growth forms in *B. attenuata*. Genome-wide association study identified 69 single nucleotide polymorphisms, putatively associated with growth form, whereas no growth form- or phenotype-specific genotypes were identified. Genomic differentiation between the two growth forms was low (Fst=0.024). Differential gene expression analysis identified 37 genes/transcripts that were differentially expressed in the two growth forms. A small heat-shock protein gene, associated with lignotuber presence, was differentially expressed in the two forms. We conclude that different fire regimes induce phenotypic polymorphism in *B. attenuata*, whereas phenotypic trait divergence involves the differential expression of a small fraction of genes that interact strongly with the disturbance regime. Thus, phenotypic plasticity among resprouters is the general strategy for surviving varying fire regimes.

Key words: *Banksia*, fire adaptation, genome-wide association study, lignotuber, phenotype, RAD-seq, resprouting, RNA-seq.
INTRODUCTION

Understanding the processes that generate and maintain biodiversity, and how these processes interact with environmental factors, are central to evolutionary biology. Phenotypic trait divergence between populations can result from contrasting responses to different environments (Nosil 2012) and is accompanied by either genetic differentiation (Smadja and Butlin 2011; Nosil and Feder 2012) and/or phenotypic plasticity (Wray 2013). Phenotypes, genomic structures and environmental factors interact closely during evolutionary divergence. Identifying the genes or genomic regions underlying the evolution of divergent phenotypes, and identifying the relative role of multiple environmental factors in shaping the pattern of phenotypic variation and genetic differentiation are important for understanding the genomic basis of adaptive divergence (Rieseberg et al. 2003; Lexer et al. 2008; Nosil et al. 2009; Nosil and Feder 2012).

Recent advances in high-throughput next-generation sequencing approaches, such as de novo genome assembly and genome re-sequencing, using restriction site-associated DNA sequencing (RAD-seq), enable the detection and genotyping of millions of single nucleotide polymorphisms (SNPs) throughout the genome (Baird et al. 2008). Statistical improvements in genome-wide association studies (GWAS), e.g., general linear modelling for structured populations (Bradbury et al. 2007), enable robust identification of alleles that are associated with a particular phenotype. GWAS require a large number of samples to be sequenced, and this has constrained studies of natural (i.e. non-commercial species) populations.

Measures such as using a reference genome, increasing the number of markers, high frequency and filtering out markers with low minor allele frequencies, can help to increase statistical power in case-control studies with small sample size, e.g. 50–100 samples (Hong and Park 2012; Wu et al. 2015). The ability to analyse expressed genes using transcriptome sequencing allows correlations between trait variation and gene expression to be identified (Wray 2013). The advantage of these approaches is the ability to identify the genetic basis for phenotypic trait divergence in organisms in their natural settings, including the presence of multiple varying environmental factors (McKown et al. 2014; Steane et al. 2017).

Fire has played a major role in shaping functional trait evolution in fire-prone regions, globally (Keeley et al. 2011; He and Lamont 2018), including Southwestern Australia (SWA), for at least 80 million years (He et al. 2011, 2016; Lamont and He 2016). Following severe crown damage by fire, resprouting occurs through new shoot growth from surviving buds or meristems that are located aerially, basally or underground (Clarke et al. 2013). Although
resprouting may occur following a broad range of disturbances, it is increasingly viewed as an adaptation to recurrent fire in fire-prone environments; indeed, resprouting may greatly enhance plant fitness in, and has long been associated with, fire-prone environments (Keeley et al. 2011; Lamont et al. 2011). Resprouting after fire enables a plant to rapidly occupy empty space, pre-empting available water, nutrients and light (Lawes and Clarke 2011; Clarke et al. 2013), and thus, plants mature more quickly than recruits from seed. Consequently, early access to pollinators and seed dispersal agents is facilitated (Lamont and Downes 2011).

Epicormic resprouting occurs via accessory buds stored in branchlets, or on main branches or the tree trunk (Burrows 2013). Such resprouting from fire-damaged stems is possible because thick bark and wood serve a heat-insulating function that protects sunken buds, ensuring that individuals rapidly recover their pre-fire dimensions (Pellegrini et al. 2017). Resprouting also occurs from buds stored in lignotubers, the swollen woody structures that arise from the axils of seedling cotyledons (Paula et al. 2016) and become buried in the soil as they develop. Lignotubers provide plants with a bank of underground buds that are insulated from high fire temperatures by the surrounding soil, bark and wood (Nobel 2001) and enable them to resprout prolifically after severe disturbances, especially fire, remove most of the aboveground biomass (Paula et al. 2016). Epicormic buds are ontogenetically identical to lignotuberous buds and differ only in their aerial location and their type of supporting structure.

Plants with lignotubers are generally shrub-like, with many similar-sized stems arising from the lignotuber, whereas plants that resprout epicormically are trees with a single dominant trunk (rarely two or three) and a few major branches. In *Banksia* (Proteaceae), a dominant group in plant-species-rich SWA, there are five known species with both lignotuberous and epicormic resprouting populations (Taylor and Hopper 1988, Lamont and Markey 1995), suggesting that these resprouting modes could be the result of highly flexible phenotypic structures.

Previous work has shown that *Banksia attenuata* has two distinctive growth forms: a tree form that resprouts epicormically, post-fire, and a shrub form that resprouts from a lignotuber. The two growth forms display consistent morphological differences, and there are no records of the coexistence of both forms within a population (Taylor and Hopper 1988; He 2014; Figure 1). The morphological divergence patterns between tree and shrub populations show significant concordance with the divergence of neutral genetic markers, and the genetic populations are clustered into two groups (tree and shrub populations), with a minor genetic differentiation between the groups of $F_{ST} = 0.27$ (He 2014). Despite the striking morphological divergence and absence of cross-pollination studies between the groups, its taxonomic status as one species has been accepted without dispute (Taylor and Hopper 1988; Mast and Thiele 2007).
Studying ecotypes that comprise morphologically distinct variants may provide insights into local adaptation, ecological speciation, and genes that determine particular phenotypes (Seehausen et al. 2014). Characterizing the genetic mechanisms and evolutionary processes that lead to phenotypic divergence and eventually promote reproductive isolation between populations is challenging. When adaptation/selection and drift occur in the same direction, disentangling the underlying genetics of traits becomes difficult. Studies that focus on divergence patterns within the genome show relatively low genomic differentiation interspersed with genes or genomic regions with significant differentiation (Hanikenne et al. 2013; Renaut et al. 2013; Sadier et al. 2014; Soria-Carrasco et al. 2014). As such, B. attenuata is an excellent species for examining the contribution of both genomic characteristics and environmental gradients to phenotypic trait divergence related to resprouting mode (i.e. growth form).

Given the clear morphological divergence between the B. attenuata tree and shrub forms (He 2014), in this study we addressed the following questions. What environmental factors (fire, climate) have driven the morphological divergence of the populations showing contrasting growth forms? Is genomic differentiation correlated with morphological divergence, and if so, what candidate genes are associated with this morphological divergence? To address these questions, we analyzed phenotypic trait variation between the two growth forms to identify which environmental factors might have driven the phenotypic divergence. We linked genotype with phenotype, in a trait association framework, using a general linear modelling approach and functionally validated the identified candidate genes using differential gene expression analysis. By combining genomic, phenotypic and environmental analyses, we sought to understand how genetic characteristics and the environment have shaped phenotypic trait divergence.

RESULTS

Morphological divergence and environmental determinants

The tree and shrub forms of B. attenuata have two distinctive morphologies (Table 1; Figure 1). The shrub form has multiple stems [12 (mean) ± 7 (SD)] arising from a lignotuber, with the largest stem diameter at 3.0 ± 0.8 (SD) cm, 50 cm above the ground and a height of 1.6 ± 0.8 m (Table 1). Shrubs possess a lignotuber up to 6 m in circumference that is 110 times larger than the widest stem (Figure 1A), and multiple stems sprout from the lignotuber after fire (Figure 1B). Individual adult trees have one major stem (rarely up to three) (Figure 1B), with an average
diameter of 23.5 ± 13.2 cm, 50 cm above the ground. Adult trees average 5.39 ± 3.14 m tall and lack a detectable lignotuber (Table 1). However, all juveniles of the tree form have a lignotuber (Figure 1D, E). In the shrub population, the number of stems increases with increasing lignotuber size (Figure 2A), whereas in the tree population, multiple stems were observed arising from the base in juveniles, but not in adult trees (Figures1E, 2B).

The multiple linear regression model analysis showed that the phenotypic differentiation in plant height, stem width, number of stems and lignotuber presence is related to average fire intervals (Table 2). Plants in regions with frequent fires (fire intervals <20 years) are shorter, with more and thinner stems than plants in areas with less frequent fires. All plants in frequently burnt regions have a lignotuber, whereas adult plants in less frequently burnt areas lack a lignotuber. Phenotypic differentiation in stem width, number of stems and lignotuber presence is not correlated with long-term average rainfall or maximum temperatures (Table 2). Plant height is correlated with long-term average rainfall: plants are shorter in low rainfall regions than in more mesic regions. There was no evidence of interaction between climatic factors and fire interval on the growth form traits.

Single nucleotide polymorphisms, population genomic structure and growth form differentiation

Using the 4,129 scaffolds with a total length of 545 Mb as the reference genome, a total of 382,287 bi-allelic single nucleotide polymorphisms (SNPs) was identified across the genome. These SNPs were present in all 29 shrubs and at least 48 (out of 50) trees. A neighbour-joining tree constructed from the total number of SNPs also separated the 79 samples into a tree group and a shrub group (Figure 2C). Genome-wide F$_{ST}$ ± SD values averaged 0.0215 ± 0.0200 (Figure 3A). There were 69 SNPs with elevated F$_{ST}$ values (0.21 ± 0.13) at 14 scaffolds, but none of the sites continuously spanned a genomic region longer than 5 kb (Figure 3A).

Excluding the divergent SNP sites, F$_{ST}$ averaged 0.0204 ± 0.0077 in the remaining sites, apparently representing genomic regions not differentiated between tree and shrub populations.

Markers associated with growth form

Among the 382,287 SNPs markers, associations between 11 markers and lignotuber presence were identified at $P < 0.001$ after Bonferroni correction. These SNP markers were scattered among the 4129 genomic scaffolds. We identified 40 markers associated with the number of stems, 18 markers with plant height, and 11 with the presence of a lignotuber, whereas none was associated with stem size (Figure 3B). Each of the 69 SNPs was exclusively associated with a
single trait. All these markers showed extensive allele sharing between the tree and shrub forms, and there were no tree or shrub-specific genotypes.

**Differentially expressed genes among juvenile shrubs and trees**

Using the FDR approach (with \( Q < 0.01 \), corresponding to \( P < 7.3 \times 10^{-6} \)), differential gene expression (DEG) analysis identified 32 unigenes/transcripts as significantly different between shrubs and trees (Figure 4). This result corresponds to 0.046% of the 68,511 unigenes and transcripts identified. Twenty unigenes/transcripts were expressed significantly less (down-regulated) in juvenile trees than in juvenile shrubs, whereas 12 genes/transcripts were up-regulated in tree juveniles (Figure 4). The 32 DEGs were grouped according to 24 Gene Ontology (GO) biological process annotations. The complete classifications are provided in Table S4. Among the 24 GO biological processes with DEGs between tree and shrub juveniles, 13 were related to stressful growing conditions, including heat, drought, salts and pathogens. Responses to karrikin, salicylic acid and jasmonic acid were also among the most significant DEGs (Figure 4). One SNP marker (out of 69), associated with lignotuber presence, was located in the intron of a gene, and this gene was expressed significantly higher in *B. attenuata* shrubs than in trees. This gene (CL7828) was classified in GO terms as a response to stress. A Blast analysis of the sequenced transcript to NCBI Genbank returned 138 hits (with E-value < 10E^{-30}), all of which encoded a 17.6 kDa class one heat-shock protein (sHDP).

**DISCUSSION**

Although strictly lignotuberous, *Banksia* species are most likely to occur in the drier environments of SWA (Lamont and Markey 1995), lignotuber presence in adult *B. attenuata* plants is not correlated with long-term average rainfall or maximum temperatures (Table 2). The seedlings of lignotuberous species were less tolerant of drought than nonsprouters, possibly because early growth resources are used for lignotuber rather than rapid taproot growth that is critical for early survival (Enright and Lamont, 1992; Canham et al. 2015). Walters et al. (2005) reported that water and nutrient availabilities did not affect the lignotuber growth and resprouting ability of some eucalypt species. It therefore seems that lignotuber formation in *B. attenuata* is not a direct adaptation to dry climates. Instead, our multivariate linear model analysis suggests that variation in fire interval has driven divergence in the resprouting mode and its associated characteristics, including plant height, stem size and number of stems. Nevertheless, stature is also positively associated with rainfall. Thus, the expected slower
growth rate under more xeric conditions would expose young plants to crown fires for a longer period, prolonging and re-enforcing the lignotuber stage when burnt (Figure 5).

Wildfires are frequent (13–22-year intervals, Enright et al. 2012) in the heathlands of SWA and are so intense that they burn the crowns of all species. Thus, *B. attenuata* is top-killed to ground level and resprouts from lignotubers. In a region with frequent fire, and where survival is water and nutrient-limited, seed availability in SWA is the major constraint on population viability, even among resprouters (Enright et al. 1998).

Lignotuberous plants are shrubby with a stature < 2.5 m, and this allows plants to allocate a greater fraction of their above-ground resources to seeds and protective structures rather than producing a large trunk (Groom and Lamont, 2011). In support, the number of cone-supporting stems increases with increasing lignotuber size in *B. attenuata* shrubs (Lamont et al. 2011). There is also a strong relationship between lignotuber size and the number of buds it stores to explain how plants with large lignotubers are more likely to survive frequent fires (Paula and Ojeda 2006).

At the mesic end of the rainfall gradient, fires are far less likely to reach the canopy (Lamont et al. 1994), and over the past 70 years (at least) tree-crown-reaching fires have occurred at an estimated 60–140-year intervals (He et al. 2016). This allows continuous growth of the main trunk, with trees recovering from fire through resprouting from the trunk and main branches rather than the base. Epicormic resprouting allows slow-growing resprouters to rapidly re-establish a broad photosynthetic surface postfire while minimising investment in attaining height (Pausus et al. 2004; Lamont et al. 2011).

Intense crown fires may be less frequent in mesic banksia woodlands and open eucalypt forests, whereas ground surface fires at lower intensity in the same habitat are much more frequent (Hassell and Dodson 2003). Although the advantages of epicormic resprouting are apparent in ground-fire systems, seedlings and saplings are vulnerable to even mild fires because their protective structures (e.g. thick bark) have not yet developed (Lawes et al. 2011). The development of a lignotuber at an early stage of *B. attenuata* growth allows vulnerable seedlings and juveniles to rapidly acquire resprouting ability after fire and other disturbances that may kill aerial shoots. Given its apparent adaptive advantages, many species in mesic banksia woodland and jarrah forest in SWA, such as *Eucalyptus marginata* (Dell 1985), *B. grandis* (Burrows 1985) and *B. menziesii* (Groom and Lamont 2011), develop a lignotuber at the seedling stage but become a tree with a main trunk by the adult stage.
Pinpointing the genes driving ecological and morphological divergence is a critical goal of molecular ecology. We re-sequenced 79 samples, and GWAS distinguished 69 SNPs putatively associated with growth form, and DEG analysis identified 32 genes/transcripts that were differentially expressed in the two growth forms. Although 69 SNPs were identified as significant at the genome-wide level, there were no growth-form-specific genotypes. Alleles were shared extensively between the two growth forms and were distinguished only at the level of allele frequency.

No matched differentiation was observed between gene expression and allele frequency. Lowry et al. (2017) suggested that the relatively low density of SNP loci could sometimes prevent the detection of trait-related loci in RAD-seq scanning, but that is unlikely to be the case in our study. The high sequencing depth in all phases of our study leads to high coverage of the reference genome, a high number of identified SNPs, and a particularly high density of SNP loci. Moreover, the overall genomic differentiation between tree and shrub forms was extremely low, with an average $F_{ST}$ of 0.021 or an average $F_{ST}$ of 0.21 for significant SNPs only. Taken together, this evidence indicates that the genetic control of the growth form is far less important than environmental factors (essentially, the fire regime).

Several studies have shown that the environment has profound effects on the phenotype through heritable epigenetic processes, such as DNA methylation (reviewed in Bird 2007). There is no mixing of growth forms in any $B$. attenuata population, suggesting stable fire regimes at the regional scale as the over-riding determining factor with a long evolutionary history: this species is estimated to have arisen 19 million years ago (He et al. 2011).

GWAS identified 11 SNPs that were associated with lignotuber presence in adult $B$. attenuata. Although it is difficult to estimate their potential role in lignotuber development, one gene may be of particular relevance. Further DEG analysis showed that this SNP is situated within the intron of a 17.8 kD small heat-shock protein (sHSP). This sHSP gene was significantly up-regulated in shrub-form juveniles. Because the juveniles of both shrub and tree forms have a lignotuber in the early growth stage, the sHSP gene is unlikely to control lignotuber development itself.

The production of high levels of heat-shock proteins can be triggered by exposure to different kinds of environmental stress conditions, such as nitrogen starvation or water deprivation, and protein up-regulation is part of the stress response (Waters 2013). The up-regulated sHSP expression in the $B$. attenuata shrub form might be related to the requirements for promoting lignotuber growth. However, the exact role of the sHSP gene in $B$. attenuata can only be speculated about until further research, such as knockout experiments, is performed.
Although stem size was identified as a morphological trait with clear differentiation between growth forms, we observed no associated SNPs that were significant at the genome-wide level. Indeed, the average cross-sectional area of a tree trunk (at 420 cm²) was 60 times that of the largest stem in the shrub form (at 7 cm²). We suggest that stem size in *B. attenuata* might predominantly be determined by limitations in resource allocation and the growing time available for each stem type, not by genetic determinants.

In shrubby *B. attenuata*, carbon and nutrient resources are allocated for the growth of the lignotuber itself, as it is a store of resources required to support post-fire bud burst and early shoot growth (Clarke et al. 2013). For example, Dell (1985) reported that phosphorus concentration in the *E. marginata* lignotuber is twice as high as in the stems. In addition, the presence of many equal stems implies that the allocation of resources to each stem will be diluted. Enright et al. (2014) noted that new stems arising spontaneously from the *B. attenuata* lignotuber during the inter-fire period would further consume available resources. However, the *B. attenuata* shrub form is subject to frequent fire that constantly kills all stems to the ground, such that the new stems do not have sufficient time to grow into a large trunk. Unlike the shrub form, the *B. attenuata* tree form would have the most resources to support the growth of one or rarely two dominant stems.

It is worth emphasising that only one of the 69 SNPs identified as associated with phenotypic traits was differentially expressed at the juvenile stage. Furthermore, for genes that were differentially expressed at the juvenile stage, only one SNP within an intron was identified as associated with any of the phenotypic traits. Gene expression is usually developmental-stage-specific and tissue-specific. One plausible reason for these results is that these SNPs may be associated with genes that are expressed only at a later stage of development, such as during post-fire recovery.

This DEG may also be triggered by a particular environment. For example, the post-fire environment, with its increased availability of nutrients, water and light (Causley et al. 2016) and reduced competition, may trigger the differential expression of certain genes, ultimately contributing to the divergence of phenotypic traits. Because the material used in our DEG analysis was from a controlled environment, we were not able to capture any such differential gene expression. Future research requires experiments in completely natural settings and including all developmental stages.

**CONCLUSIONS**
We conclude that fire interval, rather than climate, has been the most significant factor driving phenotypic divergence of the *B. attenuata* growth form. Our whole genome re-sequencing and DEG analysis revealed only a few SNP markers (0.017%) associated with morphological divergence, and none was associated with resprouting mode- or phenotype-specific genotypes. Divergence in phenotypic traits reflects allelic frequency variation and the differential expression of a small number of genes strongly controlled by the environment. Our results indicate that environmental factors, essentially fire regime, likely induce phenotypic polymorphism in *B. attenuata* and many other plant species with a similar life history and distribution.

It is clear that lignotuberous and epicormic resprouting are variations in a single regeneration trait in *B. attenuata*. Lignotubers that support many equal-sized, post-fire stems occur in an environment with more frequent crown fires, whereas trunks with epicormic resprouts exist with less frequent crown fire and are favoured in higher rainfall areas due to faster stem growth (as illustrated in Figure 5). It is likely that phenotypic plasticity of the resprouting mode is a general strategy for surviving variable fire regimes in fire-prone environments. Further work is needed to sequence a higher number of samples, analyses that annotate the *B. attenuata* genome, and DEG analysis in a completely natural setting that will likely elucidate functional genes driving this phenotypic divergence. Comparative genomic analysis of resprouters and nonsprouters would further assist in pinpointing functional genes that maintain survival and regeneration in fire-prone environments.

**MATERIALS AND METHODS**

**Study species and phenotypic measurements**

*B. attenuata* is diploid (2n = 28) and shows complete outcrossing (Scott 1980). This species has the widest distribution of all *Banksia* species in SWA, extending from Kalbarri in the northwest to the Fitzgerald River in the southeast, spanning 850 km south to north (Table S1) and constrained between the 400-mm rainfall isoline to the east and the SWA coastline to the west and south (He et al. 2016). Phenotypic measurements were taken at nine sites across the entire species range. Mature resprouting *B. attenuata* regrow rapidly after fire and regain around 80% of their pre-fire height after just 2–3 years and 100% within 5–6 years, and there-after height and size increase are very slow (Enright and Lamont 1992). These sites had not been
burnt for at least ten years, receive annual rainfall in the range 400–1000 mm, and have experienced a range of average fire intervals from ~12 years to >140 years (He et al. 2016).

At each site, 10 to 12 mature *B. attenuata* plants were sampled during a random walk, with neighbouring samples at least 30 m apart. Four phenotypic measurements were made on each plant. Height was measured from the ground to the tallest point of the plant using a clinometer or measuring tape. The number of stems >5 mm in diameter arising from the base of the plant was recorded. Stem size was measured as the perimeter of the main stem (later converted to diameter assuming the stem is circular) 50 cm above the ground as the height of stems of shrub form was generally between 80–150 cm only. The widest stem was considered as the main stem if multiple stems were present.

The size of the lignotuber was the perimeter around the base of the plant and measured by running a tape measure around all basal stems. For tree forms without a visible lignotuber, the basal perimeter was recorded as if they had a lignotuber. The presence of a lignotuber was quantified as the value of \((R_0 - R_{50})/R_{50}\), in which \(R_0\) is the radius of the basal area, assumed to be a circle, and \(R_{50}\) is the radius of the main stem measured 50 cm above the ground. A value greater than 2 was taken to indicate the presence of a lignotuber.

Lignotubers have previously been observed in juvenile tree stages among *Banksia* species, e.g., *B. grandis* (Burrows 1985), *B. menziesii* (Groom and Lamont 2011) and *B. attenuata* (this study). The presence of a lignotuber was examined on >100 juveniles (defined as <50 cm tall with no evidence of flowering) among tree populations at Jandakot Nature Reserve, 15 km S of Perth. At the same location, *B. attenuata* plants in a 50 m × 50 m plot were surveyed for main stem size (perimeter at 50 cm above the ground) and the number of stems. For the shrub form, 150 plants were measured for lignotuber size and stem number at Beekeepers Nature Reserve, 350 km north of Jandakot Nature Reserve.

Climatic data for each of the nine sites, including the two most limiting climatic factors in southwest Australian ecosystems (Groom and Lamont 2015), annual rainfall and maximum temperature, measured as the average temperature of the hottest month (February in the southern hemisphere), were taken from the nearest weather station (usually situated within 20 km of the sampling site), and data from 1960 to 2014 were retrieved from the Bureau of Meteorology Australia (www.bom.gov.au). Vegetation at all sites experienced periodic fire occurrences, and average fire intervals at each site were obtained from He et al. (2016).

**Genomic analysis of contrasting growth forms**
De novo assembly of the *Banksia attenuata* reference genome

DNA samples for reference genome sequencing were extracted from the leaves of a *B. attenuata* tree growing at the Curtin University Field Trial Area. Seven paired-end sequencing libraries were constructed with insert sizes of 170 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb and 20 kb. The libraries were then sequenced using the Illumina HiSeq2000 platform at the Beijing Genomics Institute (BGI, Shenzhen, China), yielding a total of 257.18 Gb raw data (Table S2). The raw sequences were filtered against low-quality (quality value of ≤5), adapter sequence, pair-end read overlaps, and PCR duplicates. We obtained 114.68 Gb clean data.

The genome was assembled with SOAPdenovo (Luo et al. 2012). The *de novo* assembled genome size was estimated at 661 Mb using a k-mer frequency (Genome size = K-mer_number / Peak_depth), similar to the genome size (652 Mb) of *Macadamia integrifolia*, a distantly related species in the same subfamily, Grevilleoideae (Nock et al., 2016). The final genome coverage was 77-fold. The N50 scaffold length in the assembly was 199.9 kb with a total length of 640 Gb (96.8% representation), the largest contig was >66 kb, and the largest scaffold was >1.65 Mb. The 4129 scaffolds larger than 10 kb represented a 545 Mb sequence (82.5% representation) and were used as the reference genome for further analysis (Table S3).

Restriction-site-associated DNA re-sequencing for genomic differentiation

Leaves from 8 to 10 *B. attenuata* plants were collected from the nine sites for re-sequencing using a restriction-site-associated DNA-sequencing technique (Miller et al. 2007). Genomic DNA from each sample was extracted and digested with *Eco*RI, and DNA fragments of the desired length were gel purified after electrophoresis. Adapter ligation and DNA cluster preparation were performed before the DNA fragments were subjected to sequencing in a Hiseq2000 platform at BGI. Data from multiple Illumina sequence channels were segregated by the appropriate 4–8 bp nucleotide multiplex identifier assigned to each sample. The raw data were cleaned using the following two steps: first, the adapter pollutions and index sequence in reads were deleted. Second, the reads that contained >50% low-quality bases (quality value ≤5) were removed. Thus, 79 samples (29 shrubs and 50 trees) collected across the species range and exhibiting clear morphological divergences were successfully sequenced using 100 bp pair-end RAD sequencing. For each sample, RAD-seq generated an average of 16 million reads and an average of 1.5 Gb sequence data (Table S2). The average quality score was 99.4% (call accuracy rate), with 99.1% being the lowest.
RNA sequencing for differential gene expression

To perform differential gene expression analysis of *B. attenuata* tree and shrub forms, the juvenile transcriptome was sequenced. Fruits from five trees and five shrubs were collected from each of six sites, and seeds were extracted and germinated at 15°C before transfer to pots with washed silica sand. Seedlings were grown in a growth cabinet for eight weeks (for *B. menziesii*, a congener usually co-occurring with *B. attenuata*, anatomical evidence of lignotuber development is evident by eight weeks, Mibus and Sedgley 2000). Seedlings were harvested, and RNA was extracted from the whole plant and treated with DNAase I. Because tree and shrub forms are extreme phenotypes with no intermediate phenotype, we prepared five samples for transcriptome analysis with two shrub samples and three tree samples (a third shrub sample was initially included but was discarded because of sequencing failure). Each sample contained RNA from five seedlings.

RNA sequencing was performed at BGI. Briefly, after extraction, an oligo(dT) primer was used to isolate mRNA. The mRNA was fragmented, and cDNA was synthesised using fragmented mRNA as the template. The fragments were connected to adapters after single nucleotide (adenine) addition. Suitable fragments were selected for PCR amplification. The Agilent 2100 Bio-analyser and ABI StepOnePlus Real-Time PCR System were used to quantify the sample libraries. The libraries were sequenced using Illumina HiSeq 4000 with the 100 bp pair-end method at BGI. Standard reagents for the TruSeq v3 series were used for mRNA library construction and cluster generation sequencing. The five samples were pooled into a lane to generate outputs of 60 million 100 bp single-end reads per sample. Raw reads with a ≥50% low-quality base, that were adaptor-contaminated, or that exhibited a >50% unknown base were filtered out, resulting in an average of 53.5 million clean reads (42.8–71.2) with 94.6% of reads with quality Q20 (call accuracy >99%) generated for each sample. The clean reads were first *de novo* assembled in Trinity (Haas et al. 2013) before clustering transcripts to the unigenes using TGICL (Pertea et al. 2003). A total of 68,511 unigenes with a mean length of 920 bp and a total length of 76.3 Mb were obtained (Table S2) for downstream analysis.

Data analysis

Morphological divergence and environmental determinants

The relative contributions of three environmental factors (rainfall, maximum temperature, fire interval) to morphological divergence (height, lignotuber size and number of stems) were analysed using multiple linear regression model with each trait analysed in a single model.
Analyses were implemented in PAST3.0 (Hammer, 2010). Differences were considered significant at $P < 0.05$. The Bonferroni correction was employed for multiple comparisons.

Genomic differentiation among populations of trees and shrubs

Clean RAD-seq reads were mapped to the de novo assembled B. attenuata genome (scaffolds $>10$ kb) using CLC Genomics Workbench 8.0 (CLC-GW, https://www.qiagenbioinformatics.com). Mapped reads were further locally realigned to enhance alignment near indel polymorphisms in CLC-GW before being exported as BAM files to Analysis of Next-Generation Sequencing Data (ANGSD) (Korneliussen et al. 2014) for further examination. We separated the samples into tree or shrub populations and examined genome-level genetic differentiation between the two populations. A 2D-Site Frequency Spectrum (2D-SFS) method (Nielsen et al. 2012) was used following the ANGSD tutorial (Korneliussen, http://www.popgen.dk/angsd/). Briefly, genotype likelihoods were calculated from sequence data, posterior probabilities of Sample Allele Frequency (SAF) for each site in both tree and shrub populations were estimated, and a 2D-SFS was computed. The population SAF and 2D-SFS were used to calculate per-SNP $F_{ST}$ indexes. Genome-wide differentiation between the tree and shrub population samples was assessed by sliding window calculations of $F_{ST}$ along the 4,129 scaffolds $>10$ kb using a window size of 10 kb (with an average of 7.2 SNP sites) and step size of 1 kb.

SNP-trait association analysis of identified significant markers

Variant discovery analysis was implemented in ANGSD using a maximum likelihood approach to infer major/minor SNP states based on genotype likelihoods (Korneliussen et al. 2014). For genotype calling, we used a likelihood ratio test statistic for the allele frequency based on a $w^2$ distribution and a $P$-value threshold of $1 \times 10^{-6}$. SNPs were recorded if they could be genotyped in $>95\%$ of the sampled individuals from both tree and shrub populations.

SNP-trait association analysis was performed with the software program TASSEL 5.0 (Bradbury et al. 2007) using a General Linear Model (GLM) procedure. GLM considers population structure as a fixed factor using the equation $y = x\beta + \mu$, in which $y$ is the trait value, $x$ is the marker value, $\beta$ is the matrix of parameters to be estimated, and $\mu$ are fixed cofactors (for errors and false positives caused by the population substructure; Bradbury et al. 2007). Population structure was estimated using principal component analysis following Zhao et al. (2007). The multiple-testing threshold level was set at $2.6 \times 10^{-9}$, equivalent to a Bonferroni-corrected $P = 0.001$ based on the number of SNPs analysed (Matthews and Foulkes, 2015). A neighbour-joining dendrogram containing the 79 tree and shrub samples was also generated.
using all the called SNPs with TASSEL to visualise possible phylogenetic relationships between the samples.

**Comparative analysis of juvenile tree and shrub transcriptomes**

Clean reads from each sample were mapped to *de novo* assembled unigenes, and the analysis of differentially expressed genes (DEGs) was performed using CLC-GW. The number of reads mapped per transcript was converted to reads per kilobase per million reads (RPKM, Mortazavi et al. 2008) to standardise expression to the length of the transcript and the depth of sequencing. DEGs between tree and shrub juveniles were determined with a false discovery rate (FDR) of <0.01, corresponding to a *P*-value <1 × 10^{-5}. Blast2GO (Conesa et al. 2005) was used to functionally annotate the DEGs between trees and shrubs. We used an FDR <0.01 for each *P*-value to identify significantly enriched biological processes. Finally, significant DEGs were aligned to contigs containing SNPs that were identified as significant in the above GWAS. Contigs containing significant SNPs were assembled into a reference genome, and significant DEGs were then aligned to the reference sequence with a local alignment option in CLC-GW. Because the expressed gene sequences encode only exomes, the alignment method is more appropriate than Blast for mapping the expressed gene in the genome. The mapped genes were manually examined for 1) significant SNP site coverage; 2) if no SNP coverage was observed, whether the significant SNP was present within 700 bp (based on one SNP discovered every 1.4 kilobases on average) at either side of the mapped gene. Either of the above two criteria identified possible genes that were differentially expressed in tree and shrub juveniles and contained significant SNPs at the genome level.

**ACKNOWLEDGMENTS**

This study was supported by the Australian Research Council (DP130013029). The Beijing Genomics Institute undertook the DNA sequencing; William Fowler assisted with fieldwork. The authors declare no conflict of interest.

**AUTHOR CONTRIBUTIONS**

T.H. and N.E. designed the experiments. H.A. and T.H. performed the experiments. T.H., B.L., and W.S. analyzed the data. T.H. and B.L. wrote the article.

**REFERENCES**

Burrows ND (1985) Reducing the abundance of *Banksia grandis* in the jarrah forest by the use of controlled fire. *Aust Forestry* 48: 63–70
Enright NJ, Keith DA, Clarke MF, Miller BP (2012) Fire regimes in Australian sclerophyllous shrubby ecosystems: Heathlands, heathy woodlands and mallee woodlands. In Flammable...


Groom PK, Lamont BB (2015) Plant Life of Southwestern Australia: Adaptations for Survival (De Gruyter Open, Warsaw, Poland)


sunflowers facilitated by hybridization. Science 301: 1211–1216
SUPPORTING INFORMATION

Table S1. GPS locations of the nine studied sites (the location of the first plant measured)
Table S2. Data generated for *de novo* assembly of *Banksia attenuata* genome
Table S3. Summary of genomic data for *Banksia attenuata*
Table S4. Complete list of differentially expressed genes different between shrubs and trees

Figure legends

Figure 1. Morphology of shrub and tree populations of *Banksia attenuata*
(A) Shrub form. (B) Tree form. (C) Post-fire resprouting from a lignotuber, dead stems are 1 m tall. (D) (1) Juvenile in the tree population, 30 cm tall and about three years old. (E) (2) Multi-stemmed, lignotuberous juvenile; (3) Dominant-stemmed sapling, 2 m tall, and a single-stemmed tree (4), 5 m tall.

Figure 2. Presence of lignotubers among tree and shrub populations of *B. attenuata*
(A) Lignotuber size (perimeter) and number of stems in the shrub population. (B) Stem diameter and number of stems in the tree population. (C) Neighbour-joining tree constructed using all detected SNPs from the analysed *B. attenuata*, showing the two distinct groups, corresponding to the two growth forms. Each bar represents samples from the same population.

Figure 3. Genomic differentiation between the tree and shrub growth forms and association of SNPs and phenotypic characters
(A) Genomic differentiation along scaffolds between the two growth forms as calculated through sliding window. (B) significant SNPs along scaffolds identified to be associated with
each of the four phenotypic characters using GWAS. Red dotted line indicates the $P$-value threshold.

**Figure 4. Differential pattern of gene expression in tree and shrub form seedlings**
Left: Differentially expressed genes between the two growth forms; red dots indicate significant deferential expressed genes between the two growth forms. Right: the GO (Gene Ontology) terms of the significant deferential expressed genes between the two growth forms.

**Figure 5. Conceptual model of processes involved in, and maintenance of the two growth forms in Banksia attenuata, showing the determining role of fire frequency in the habitats**

**Table 1. Morphological measurements for Banksia attenuata tree and shrub populations at nine sites**

<table>
<thead>
<tr>
<th>Site</th>
<th>Height (m)</th>
<th>Stem size (cm)</th>
<th>No. stems</th>
<th>Lignotuber $^b$</th>
<th>Annual rainfall (mm)</th>
<th>T-max $^c$ ($^{\circ}$C)</th>
<th>FI (year)$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalbarri NP</td>
<td>1.4±0.1</td>
<td>9.2±1.6</td>
<td>13.1±1.7</td>
<td>40.2±12.0</td>
<td>345</td>
<td>34.3</td>
<td>15</td>
</tr>
<tr>
<td>Beekeeper NR</td>
<td>1.4±0.2</td>
<td>9.5±1.8</td>
<td>11.7±7.1</td>
<td>36.8±19.8</td>
<td>491</td>
<td>36.4</td>
<td>13</td>
</tr>
<tr>
<td>Eneabba</td>
<td>0.9±0.3</td>
<td>9.4±3.9</td>
<td>10.8±7.3</td>
<td>40.4±27.0</td>
<td>510</td>
<td>34.8</td>
<td>13</td>
</tr>
<tr>
<td>Fitzgerald NP</td>
<td>2.8±0.5</td>
<td>46.5±11.2</td>
<td>2.1±1.5</td>
<td>1.5±0.2</td>
<td>385</td>
<td>28.7</td>
<td>95</td>
</tr>
<tr>
<td>Arthur River</td>
<td>7.0±1.3</td>
<td>122.3±36.3</td>
<td>1.5±0.7</td>
<td>1.1±0.1</td>
<td>431</td>
<td>31.1</td>
<td>65</td>
</tr>
<tr>
<td>Grace town</td>
<td>7.5±2.7</td>
<td>58.5±31.2</td>
<td>1.0±0.1</td>
<td>1.5±0.2</td>
<td>1001</td>
<td>27.2</td>
<td>81</td>
</tr>
<tr>
<td>Brunswick</td>
<td>9.3±3.1</td>
<td>114.3±38.3</td>
<td>1.1±0.3</td>
<td>1.2±0.2</td>
<td>987</td>
<td>33.1</td>
<td>100</td>
</tr>
<tr>
<td>Goomalling</td>
<td>3.5±0.5</td>
<td>59.8±20.9</td>
<td>1.7±0.8</td>
<td>1.3±0.2</td>
<td>366</td>
<td>34.9</td>
<td>63</td>
</tr>
<tr>
<td>Yanchep NP</td>
<td>2.5±0.8</td>
<td>42.1±10.8</td>
<td>1.2±0.6</td>
<td>1.4±0.1</td>
<td>739</td>
<td>33.3</td>
<td>40</td>
</tr>
</tbody>
</table>

a: trunk circumference at 50 cm above ground; for individuals with multiple stems, the biggest was measured; b, quantified used the formula described in the Methods; a value greater than 2
was regarded as presence of a lignotuber; c: data extracted from He et al. (2016). NP = National Park, NR = Nature Reserve, T-max = mean temperature of warmest month, FI = Fire interval

<table>
<thead>
<tr>
<th>Morphology</th>
<th>Environmental factor</th>
<th>$P$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Annual rainfall</td>
<td>0.0000</td>
<td>0.366 +</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature</td>
<td>0.2707</td>
<td>0.216 -</td>
</tr>
<tr>
<td></td>
<td>Fire interval</td>
<td>0.0000</td>
<td>0.489 -</td>
</tr>
<tr>
<td>Stem size</td>
<td>Annual rainfall</td>
<td>0.4787</td>
<td>0.103 +</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature</td>
<td>0.0163</td>
<td>0.118 -</td>
</tr>
<tr>
<td></td>
<td>Fire interval</td>
<td>0.0000</td>
<td>0.440 -</td>
</tr>
<tr>
<td>Number of stems</td>
<td>Rainfall</td>
<td>0.3167</td>
<td>0.131 -</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature</td>
<td>0.9551</td>
<td>0.232 +</td>
</tr>
<tr>
<td></td>
<td>Fire interval</td>
<td>0.0000</td>
<td>0.457 +</td>
</tr>
<tr>
<td>Lignotuber size</td>
<td>Annual rainfall</td>
<td>0.8448</td>
<td>0.109 -</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature</td>
<td>0.9681</td>
<td>0.262 +</td>
</tr>
<tr>
<td></td>
<td>Fire interval</td>
<td>0.0000</td>
<td>0.527 +</td>
</tr>
<tr>
<td>PC1</td>
<td>Annual rainfall</td>
<td>0.4616</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>Maximum temperature</td>
<td>0.0211</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>Fire interval</td>
<td>0.0000</td>
<td>0.501</td>
</tr>
</tbody>
</table>

Rainfall is the average annual rainfall 1960–2014 and maximum temperature is the average temperature of the hottest month (February) 1960–2014. PC1 is the first principal component of four morphological traits. Significant factors are in bold. Sign beside $R^2$ is the direction of the correlation.
Figure 1
Figure. 2
Figure 3

A: Genetic differentiation between tree and shrub populations through sliding window analysis
window size: 10 kb, step size 1 kb.

B: Association between SNPs and phenotype

- Lignotuber
- Number of stems
- Height
- Stem size

Genomic position
Figure. 4

Figure. 5