A systems genetics approach to understanding the control of natural variation of leaf morphology in European aspen

Kathryn M Robinson[+,1], Niklas Mähler[+,2], Barbara K Terebieniec[1], Torgeir R Hvidsten[1,2], Nathaniel R Street[*1]

[*] Corresponding author: Nathaniel R Street: nathaniel.street@umu.se
[+] Equal contribution

[1] Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, 901 87 Umeå, Sweden
[2] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1430 Ås, Norway

Abstract

Background: A central aim of biological investigation is to understand how genomes encode information controlling emergent, complex phenotypes and the genetic architecture underlying natural variation of such traits among individuals. There are still relatively few studies exploring the genetic architecture of complex traits in natural populations with no consensus understanding of how genetic architecture is linked to the evolutionary history of a trait. Here, we focused on natural variation in leaf size and shape in a natural population of European aspen (*Populus tremula*) sampled across the distribution range of Sweden. Leaf morphology traits show no evidence of clinal variation of population differentiation, appearing to be selectively neutral. As such they serve as a useful contrast to studies performed on highly adaptive traits, such as the timing of autumn bud set.

Results: We assayed leaf size and shape variation in replicated common garden experiments of the Swedish Aspen (SwAsp) collection, finding no correlation between leaf shape and a range of climatic, geographic and biological traits and no evidence of population differentiation. We utilised a collection of genome-wide SNPs to perform a Genome-Wide Association Study (GWAS), identifying overlapping sets of SNPs from the replicated gardens/years in order to reduce false positives. These results were integrated with RNA-Sequencing data assaying gene expression in the SwAsp collection that has previously been used to map expression QTL (eQTL). Using these data, we identified SNPs associated to leaf physiognomy phenotypes, to gene expression (eQTL) and correlations between phenotype and gene expression levels.

Conclusions: The genetic architecture underlying natural variation in leaf morphology traits within the study population was in agreement with the infinitesimal model, with individual SNPs explaining little of the heritable trait variation. There was no evidence of correlation between gene expression and the phenotypic traits considered, although overlapping sets of morphological and expression associated SNPs were identified. We conclude that variation among genotypes in leaf morphology is controlled by the effect of many SNPs, each of small effect, each of which results in a small-scale modulation of expression patterns contributing to the control of the leaf developmental program.

Keywords: Leaf shape; RNA-Sequencing; transcriptomics; Populus tremula; association mapping, natural variation, candidate gene

Introduction

For many years the accepted dogma was that variation between species, and genetically controlled aspects of variation among individuals of the same species, resulted largely from non-synonymous changes in the protein coding sequence of genes: Resultant changes in protein structure/function then perpetuated a change in phenotype. Our understanding of the causal factors underlying natural variation and speciation has been revolutionised during the past decade and, most recently, is being further refined by discoveries arising from the application of high throughput sequencing approaches. For example, genomics has provided extensive evidence that divergence in regulation and expression network structure are key components of both within species variation and divergence between species [1–8]. This has required a change in emphasis from the identification of causal polymorphisms within the protein coding regions of genes to that of identifying sequence variations that modify gene regulation and expression. The ability to determine whether the source of such regulatory diversity derives from polymorphisms lying in *cis* (local effects) or in *trans* (distant effects) additionally enhances our understanding of the genetic architecture of gene expression diversity [4, 9].

While reductionist molecular biology approaches have taught us much about the function and role of numerous individual genes, we still know relatively little about the mechanisms underlying natural variation and how interacting networks of genes result in the emergent properties of phenotypes [9]. Complex polymorphic traits are not the result of genes acting independently but rather are emergent properties of a polygenic, dynamic system of interactions among genes and between genes and the environment [10]. While genetic approaches have provided insight into the genetic architecture of complex traits, knowledge of the causal genetic polymorphisms has remained limited. For example it is not known if the majority of causative polymorphisms lie in protein-coding, promoter, intron or inter-genetic regions of the genome, how often the control of polymorphic traits is determined by *cis* (*i.e.* proximal) or *trans* (*i.e.* distal) effects and how these patterns relate to selection and adaptive trait variation.

Partly driven by the above questions and by the availability of new technologies, the previous decade saw an explosion of interest in genome-wide association studies (GWAS), particularly in the field of human medical research. Such studies aim to identify causative genetic polymorphisms contributing to the control of complex quantitative phenotypes (for example height or obesity). Results from the numerous studies performed have, perhaps, been less insightful than hoped [11]. For the majority of traits with moderate to high heritability considered to date, potentially causative SNPs explain little of the phenotypic diversity that exists and often lack biological interpretation [12–14], leading many to ask where the 'missing heritability' lies [15]. As a result, attention is shifting towards approaches that integrate multiple forms that can both identify functionally important SNPs and provide biological context or insight as to their mode of action. These approaches also represent a means to minimise wasted effort on false-positives, which can be particularly important considering the substantial effort invested in functional validation of identified candidates. Integrative approaches include various combination of the relationship between SNPs and eQTL, differentially expressed genes and constructed expression network structure, phenotypic trait correlations and association mapping or QTL results (for example [16–20]).

Systems genetics formalises such integrative approaches as the study of systems biology within a population genetics context [9]. As an analogy, GWAS provides a two-dimensional view of a system: Adding gene expression data from the same individuals transforms this view into a three-dimensional

one, allowing previously hidden properties to be seen. In contrast to many previous systems biology approaches where, for example, gene expression networks are constructed from a diverse set of experimental conditions, systems genetics profiles expression among individuals of a population. An expression network is then constructed based on the correlation structure of expression variation across the population. In this way, hubs in the network structure represent key components determining variation within the population. Two key aspects of the systems genetics approach are that it places emphasis on the fact that complex traits are not the result of genes acting independently but rather are emergent properties of a dynamic system of interactions between genes and the environment and that the link between phenotype, expression and genomic variants can provide functional insight when identified variants lie outside of protein coding regions [10].

As model systems for genetic studies, plants have a number of distinct advantages over animals: Clonal propagation results in the ability to precisely calculate heritability, phenotypic plasticity, genotype and environment effects and their interaction (GxE) within a single generation [15]. To date the plant field has focused largely on candidate gene based association studies. A major limitation of such an approach is that only genes for which there is pre-existing knowledge are considered. An assumption is also made that those genes are involved in the determination of natural variation. Additionally, a functional candidate gene approach is typically combined with a screen for SNPs within coding regions. This is counterintuitive as the functional evidence used to identify candidate genes is often the presence of phenotypic variation associated with differences in gene expression – and expression differences are rarely caused by SNPs within coding regions. A substantial risk of such an approach is falsely concluding that a gene does not play a role in controlling trait variation simply because no SNPs within the coding region associate with the phenotype of interest. This risk is especially high in species with rapid decay of linkage.

European aspens have been shown to contain extremely high levels of genetic variation [21–23], to have no significant population structure ([24]; Wang *et al.* 2016 in prep) and to be suitable for high-resolution association mapping [25]. *P. tremula* can also be grown in tissue culture and is amenable to genetic transformation, meaning that functional confirmation can be generated in the genetic background studied. In this context, variation in leaf shape of European aspen represents an excellent model system as we have previously shown that many leaf physiognomy traits have high levels of heritable variation [26]. Leaves are the direct energy source sustaining the majority of complex life on earth. Humans interact with and recognise plant species largely through the shape of their leaves and, as such, leaf shape forms an important component of the relationship we share with the living world. Indeed, leaf shape was historically one of the key features used by Linnaeus and others to classify and identify species and long-term historical changes in leaf shape recorded in the fossil record provide insight into historical climatic conditions, subsequently allowing extrapolation of past trends to future changes in response to climate change. Leaf shape varies distinctly both between and, often, among species with some identifiable global trends such as the narrowing and more defined serration of leaves toward latitudinal extremes [27].

Here, we have taken a systems genetics approach using the Swedish Aspen collection [28] to investigate variation in a number of traits associated with leaf shape and size. These traits were selected to serve as a contrast to other highly adaptive traits that we have also been investigating (Wang *et al.* 2016 in prep) as leaf shape shows no evidence of clinal variation and has likely not been a target of positive selection. We have integrated population-wide genomic resequencing and RNA-Sequencing data together with morphological phenotypic traits to explore the genetic architecture of leaf shape variation.

Results

Phenotypes

We sampled mature, pre-formed leaves in the Swedish Aspen (SwAsp) collection at two common garden experiments in the north (Sävar, Västerbotten) and south (Ekebo, Skåne) of Sweden in two years (2008 and 2011). There was considerable variation in leaf shape represented among the SwAsp genotypes (Figure 1A) and we used a digital image analysis method [26] to measure a number of traits indicative of leaf physiognomy (Figure 1B). We calculated clonal repeatability as an upper-bound estimate of broad sense heritability (H^2), which was relatively high for all traits and for which shape traits (including circularity and measures of indent size and number) had distinctly higher values than those related to size (measures of length, width and area), accounting for the bimodal distribution observed in Figure 1C (see supplementary file S1 for details of all traits). We calculated Q_{ST}, which revealed no evidence of population differentiation for any of the traits considered (Figure 1D, supplementary file S1). As many of the dimension traits have a high degree of redundancy, we performed a dimension reduction using principle component analysis (PCA), for which PC1 was related primarily to size variation with PC2 and PC3 relating to components of shape (Figure 1E). ANOVA tests comparing traits and PCs in the two gardens/years revealed a variable degree of environmental variance (supplementary table S1), with size traits and PC1 having a greater variation than shape traits and PCs (consider the F values in sheet 1 of supplementary file S1). Although significant GxE was observed for a number of both size and shape traits, the percentage of variance explained by GxE (as indicated by F values) was considerably smaller for the shape related traits. As such, leaf shape is under tight genetic control and is a relatively invariant feature of a genotype. This is confirmed by the higher correlation between shape trait values for the two gardens/years than for size related traits (Figure 1F).

Genome wide association mapping

We performed a genome wide association (GWA) analysis to identify links between single nucleotide polymorphism (SNP) variants and indent width, leaf area, and leaf circularity for all of the four datasets. A total of 4.5 million SNPs (detailed in Wang et al. 2016, In prep) were associated to the phenotypes separately for each garden and year. After multiple testing correction and combining all significant associations from all four datasets, 65 SNPs were significant for indent width, 39 for leaf area and no SNPs were significant for circularity at a 5% false discovery rate (FDR; Figure 2). The significant SNPs for indent width had a wide range of contexts from coding SNPs to non-coding and intergenic SNPs, but judging from their location, many of them are likely to be linked. 44 of the 65 indent width SNPs were located within 2 kbp of a total of 15 genes, and all but one of the SNPs with a significant association to leaf area were located within 2 kbp of 5 genes. Neither of these gene sets were enriched for any GO terms.

To identify a set of the most consistent SNPs in order to minimise false positives, we selected the top 1000 SNPs (ranked by adjusted p-value but disregarding significance) for each trait and garden/year (figure 2). We refer to these SNPs as phenotypic trait associated SNPs or simply pSNPs. The genomic context of pSNPs identified in at least two gardens/years are shown in Figure 3 We also identified the set of genes associated with the pSNPs (genes located within 2kb of the SNP) present in at least two of the gardens/years (Materials and methods): 684 SNPs were associated with 252 genes for circularity (enriched in inorganic anion transport, GO:0015698, $p = 0.005$), 715 SNPs with 236 genes for indent

width (enriched in hexose metabolic process, GO:0019318, p = 0.007), and 66 SNPs with 43 genes for leaf area (enriched in sucrose metabolic process, GO:0031324, p = 0.0003).

Gene expression

After filtering gene expression values on variance and adjusting for hidden confounders (Materials and methods), correlations were computed between the expression of 22,306 genes and the three morphological traits for each garden and year. After correcting for multiple testing, no genes were significantly correlated to any of the traits at 5% FDR (Figure 4). We therefore investigated the top 1000 genes from each garden/year (ranked by adjusted p-value but disregarding significance) and refer to these as expression-phenotype correlated genes or epGenes. Overlaps between the epGenes from each garden/year were relatively high (Figure 4). A GO enrichment analysis showed that the intersection of genes correlating with circularity was enriched for "ATP biosynthetic process" (GO:0006754; 31 genes, p = 0.008), intersecting genes correlating with indent width were enriched for "cellular protein localization" (GO:0034613; 7 genes, p = 0.0018), while no significant GO terms were found for leaf area.

Gene set enrichment analysis (GSEA)

Since no individual genes displayed expression that correlated significantly with any of the phenotypes, we employed gene set enrichment analysis (GSEA, [29]) to test whether any *sets of genes* were significantly enriched at the extreme ends of the gene list, *i.e.* the list of all 22,306 genes sorted by correlation to a phenotype (Materials and methods).

For each of the three phenotypes, we tested two types of gene sets (Tables 1-3 for circularity, indent width and leaf area, respectively). Firstly, we tested genes associated with the pSNPs that were found in all, or at least two, of the gardens/years (see Venn diagrams in Figure 2). The only significant set from this analysis was the genes associated with the 98 pSNPs discovered for circularity in all gardens/years (p = 0.027). Secondly, we tested genes with a common functional role (Gene Ontology annotation). Here we found several significant associations including GO terms related to amino acid biosynthesis and transport, as well as DNA repair and carbohydrate metabolism

Data integration

We combined data on pSNPs, epGenes, and eQTLs (referred to as pairs of eSNPs and eGenes, see Paper III) to identify genes and genetic variants that were associated with leaf phenotypes. For each phenotype, we first intersected genes associated with pSNPs identified in at least two gardens/years, epGenes identified in at least two gardens/years and eGenes associated with significant eSNPs in the eQTL mapping. This resulted in three genes each for circularity and indent width, and none for leaf area. The three genes for circularity were Potra000998g08306, Potra001379g11776 and Potra009203g26307, annotated as disease resistance protein, synthase mitochondrial F1 complex assembly factor, and CASP-like protein, respectively. The three genes associated with indent width were Potra000351g01289, Potra163617g27107 and Potra000727g05700, all of which are unannotated. Potra163617g27107 does however have an ortholog in *Arabidopsis thaliana* that is annotated as being part of the mitochondrial outer membrane translocase complex (GO:0005742).

Next, we intersected the GSEA-significant GO categories in Tables 1-3 with the pSNPs and eQTLs. This revealed that several functional categories significantly associated to phenotypes through

expression correlation also contained many eGenes. However, they contained very few genes with pSNPs and there was almost no overlap between these pSNPs and eSNPs.

Discussion

For more than a century there has been much unresolved and speculative discussion as to the evolutionary significance, if any, of leaf shape *per se* and of natural variation of leaf shape within populations [30, 31]. While global and historical trends in leaf shape have been identified and attributed to climatic conditions [27], there remains a rather spectacular paucity of evidence linking standing leaf shape variation to aspects of plant fitness [30]. In the present study we identified considerable genetic control of leaf physiognomy traits (Figure 1), in particular traits associated with leaf shape – indeed these are the highest values we have observed for any of the numerous phenotypes we have considered to date, regardless of whether or not those traits have evidence of being under positive selection ([32], Wang *et al,* 2016 in prep). There was considerable observable variation in leaf shape both within local populations and within the distribution range of aspen across Sweden, with no identified clinal trends or evidence of population differentiation (Figure 1, supplementary file S1). Such patterns of variation may indicate that variation in the trait is selectively neutral or that balancing selection is acting to maintain within population variation. Here, we identified very few significant associations between SNPs and phenotypic traits, with no significant associations being identified in more than one garden/year for any of the traits considered. Even when identifying overlap without considering significance but taking the top 1000 SNP associations per garden/year per trait (pSNPs, Figure 2), there was still relatively little overlap, although there were greater numbers of SNPs in common among two or more gardens/years for the traits with higher H^2. Although these overlaps were low, GO over-representation tests identified significant enrichment of categories for genes located within 2kb of the pSNPs, indicating that there was biologically functional meaning to the sets of genes, albeit without readily explainable functional interpretation. Taking the total variance explained for the sets of overlapping SNPs (data not shown) reveals that very little of the total genetically controlled variance was explained by identified SNP associations. If one assumes that SNPs of high effect size would have been detected as significant, this result suggests that there remain a very large number of SNPs of small effect size that, in combination, control variation in leaf size and shape – a finding in support of the infinitesimal model. This is an area where there is little consensus in the literature, with various studies reporting genetic architectures for complex traits spanning a range from few SNPs of large effect to many SNPs of small effect size. One likely explanation for these contrasting results is that genetic architecture contrasts depending on the extent to which a trait has been the target of positive selection, with strongly adaptive traits (such as bud flush for aspen within Sweden) more likely to be explained by a small number of SNPs of large effect (Wang *et al.* 2016, In prep). The case of leaf physiognomy in aspen appears to be far more similar to that of human height, for which there appears to be a vast number of small effect SNPs underlying the height variation among individuals within a population.

For the three traits considered, the greatest proportion of associated SNPs identified in at least two gardens/years were located in UTR regions and the fewest in intergenic regions (Figure 3). The distributions of the proportion of SNPs within exons, introns and flanking regions were less consistent between the three traits, although with so few SNPs considered it is hard to interpret whether these differences are meaningful. The presence of the largest proportion of SNPs within UTRs is likely indicative of these acting to modify expression, suggesting that gene expression variation should be associated with variation in these traits. We utilised a resource profiling gene expression from winter

buds undergoing induced spring bud flush in controlled conditions in an attempt to identify gene expression variation associated with trait variance. As for pSNPs, there were relatively few consistent correlations present between gene expression levels and phenotypic traits (Figure 4), none of which were significant. Similar to pSNPs, a greater degree of overlap among gardens/years was observed for the more heritable traits (circularity and indent width) compared to leaf area; when considering the top 1000 expression-phenotype correlated genes (epGenes). In the case of circularity and indent width, significant over-representations for GO categories was present, suggesting biologically relevant functional links for these genes. In contrast, no such signal was detected for leaf area, which is likely due to the lower heritability and greater degree of environmental variability for leaf shape, possibly suggesting that different mechanisms, and therefore sets of genes, influenced leaf area at the two gardens and in the two years.

The power of systems genetics lies in the integration of expression, eQTL and phenotypic GWAS results. In our case such integration proved to be of minimal value due to the low number of trait-SNP and trait-expression associations detected. Of the three genes linked for all data types (overlap between genes close to pSNPs, epGenes and eGenes) for circularity and indent width, none of these are known to function in the control of leaf shape or to have any role in leaf development. However, these do represent our best leads for further downstream investigation.

The infinitesimal model presents a number of analysis challenges. The first is that it can be extremely challenging to identify statistically significant associations between causal SNPs (true positives) and phenotypes, especially in cases such as ours where the millions of SNPs considered creates something of a multiple testing nightmare. Also, applying methods that employ a one gene or one SNP at a time strategy to explain traits will likely prove futile. A more realistic approach is to utilise gene set enrichment analysis (GSEA, [29]), where one can identify sets of genes that together correlate significantly with a trait, despite there being no significant individual genes. Here, we utilised the list of genes sorted by expression correlation to traits to test whether gene sets (1) located close to discovered pSNPs (i.e. top 1000 SNPs identified in at least two gardens/years) or (2) with a common functional role (*i.e.* genes annotated with the same Gene Ontology category) displayed such properties (Tables 1-3). The former (1) identified only one significant association between circularity and the 98 pSNPs discovered for circularity in all gardens/years (Figure 2). The latter (2) revealed several functional categories with significant associations to traits. Moreover, several of these categories included high numbers of genes with associated eQTLs (eGenes). However, the same categories contained very few genes located close to phenotypic trait associated pSNPs, and very few of these pSNPs were in turn associated with eQTLs (i.e. few pSNPs were also eSNPs). Thus, although we identified some gene sets for which expression were significantly associated with traits and that many of these genes also had mapped eQTL associations, there was a very low correspondence between gene expression and genome sequence driven discoveries. In part, such low correspondence will be due to the fact that each SNP affecting gene expression is of small effect size (Paper III) and that expression variation of an individual gene contributing to the control of the phenotypic traits also explains only a small fraction of the total phenotypic variation. Another important factor relates to the fact that we have only a single snapshot measure of gene expression and that an eQTL can only be identified for this snapshot: another snapshot measure would be expected to identify additional SNPs associated with expression variance.

It may seem somewhat paradoxical that pSNP association results suggest that gene expression is likely to be the primary mechanism driving leaf shape variation (inferred from genomics contexct, Figure 3) but that we identified so few correlation links between expression and the traits, and furthermore that

so few of the pSNPs were eSNPs. However, there are a number of possible explanations for this lack of correspondence, some of which are already alluded to above. Although the molecular mechanisms that give rise to leaf shape variation remain entirely unknown [30], leaf development is a temporal process with many temporally separated components, variation in any of which will contribute to variation in final leaf form. As such, a single snapshot of gene expression during this developmental program will be insufficient to capture all relevant links between gene expression and final leaf form. To overcome this limitation would require a time series sampling strategy, with samples being collected at multiple developmentally equivalent stages for all individuals in the population and with eQTL mapping being performed for the snapshot measures of expression of all genes at each time point as well as for expression trajectories – a strategy that is currently intractable (or at least not fundable). In the current case of profiling expression from leaves that are produced and developmentally arrested in winter buds there are added complications. It is very likely that the period before bud set comprises an important developmental period where much of the pattern formation leading to final leaf shape is occurring. However, as genotypes set bud at different times and as climatic and environmental conditions vary considerably during the period in which bud set occurs, collecting developmentally equivalent samples of all genotypes would be extremely challenging. Before bud set the problem is also compounded by the fact that within a single bud the multiple leaves produced will be at different developmental stages. As such there would be a very high chance that the averaged expression snapshot that would be obtained from extracting RNA from a whole bud would negate any meaningful developmental expression signature (the Simpson's paradox). Even after bud set, when all leaves are arrested at an equivalent developmental stage, similar problems exist at the leaf level as leaf development also varies spatially, for example with cell production being more prolific and continuing for longer at the leaf base than at the tip. Here again, a single sample from an entire leaf will potentially mask or negate the ability to associate expression variation to phenotypic traits. In both cases, spatially resolved expression profiling, for example using laser capture microdissection, would offer a solution although performing such an experiment for replicated samples all individuals in a population would be a daunting undertaking and would likely suffer a multitude of confounding technical factors. New or improved sampling techniques and expression profiling approaches combined with falling costs of generating expression data will help to overcome these limitations or to make more comprehensive experimental designs feasible within the future, however they will not overcome the fact that the signature of SNP to expression to phenotype will be weak and hard to detect for any trait following the infinitesimal model.

Our results have shown that variation in the shape of pre-formed leaves in European aspen is controlled by numerous SNPs each of very small effect. These SNPs were primarily located in UTR regions of genes, suggesting that they induce variation in leaf form through modulation of gene expression patterns. However, very few of these coincided with eQTL and there was no significant correlation detected between a snapshot of gene expression during spring bud flush and the phenotypic traits considered.

Our findings highlight the challenges faced when employing a systems genetics strategy, especially for traits controlled according to the infinitesimal model. As such, although the approach holds great potential for providing functional insight to the link between SNPs and phenotypic trait variation, the biological characteristics of the study system may present severe limitations to this potential.

Materials and Methods

Leaf shape phenotyping

Leaf size and shape parameters were measured in a natural population of *Populus tremula*, the Swedish Aspen (SwAsp) collection, growing in common gardens at Sävar, northern Sweden (63.9°N, 20.5°E) and Ekebo, southern Sweden (55.9°N, 13.1°E). The common garden trials comprised of natural (wild-growing) aspen genotypes collected in 2003 across ten latitudinal degrees, which were cloned and planted in 2004 in a randomised block design in each garden [28]. Leaf samples were harvested in Sävar on 14 July 2008 and 28 June 2011 and in Ekebo on 18 July 2008 and 4 August 2011, when leaves were fully expanded and mature, but prior to bud set and before the occurrence of substantial damage due to herbivory or the presence of fungal rust infection. Ten undamaged leaves per replicate tree were sampled randomly across the canopy, avoiding leaves from the first or last leaf in a leaf cohort originating from a single bud. In total, in 430, 444, 326 and 393 trees were sampled in Ekebo 2008, Ekebo 2011, Sävar 2008 and Sävar 2011 respectively, comprising between 1 and 8 (median = 3) clonal replicates. One hundred and thirteen genotypes were sampled in both years in Ekebo and in 2011 in Sävar, and 111 genotypes were sampled in 2008 in Sävar. Leaves were stored at 4° - 8°C immediately after harvest. Petioles were removed at the leaf base and the sample of ten leaves per tree was scanned in colour at 300 dpi using with a CanoScan 4400F. A 5x4cm Post-it note was scanned as a scale image. The resulting images were analysed using LAMINA [26] to obtain leaf size and shape metrics (see supplementary file S2 for a list and descriptions). Median values of the ten leaves per tree were calculated for each leaf size and shape metric. Principal components analysis (PCA) was employed using the prcomp function within the R programming environment [33] for the size and the shape trait sets separately and in combination (see supplementary file S2 for the classification of obtained traits as either size or shape related). The first three principal components for leaf shape (PC1, PC2 and PC3) were used as unique leaf shape phenotypes and the first principal component for leaf size metrics (PC1) was used as a summary phenotype of leaf size. The first three components of the entire data set (Size and shape PC1, 2 and 3) were used to summarise all leaf phenotypes into three reduced descriptors. The PCA loadings are provided in supplementary file S2. For genome-wide association mapping, median values of each leaf shape and size phenotype were calculated for each aspen genotype for which there were three or more clonal replicates.

Statistical analyses

All statistical analyses were conducted in R. Phenotypic data were examined for homogeneity of variance. No data transformations were required to meet the assumptions of a normal distribution. Pearson correlations were used for all phenotypic correlations calculated.

We calculate clonal repeatability (R) and used this to provide an upper-bound estimate or broad sense heritability (H^2 – see materials and methods). We refer to this trait as H^2 rather than R as this probably allows a more intuitive interpretation for readers, however we not that the two are not the same (see (Dohm, 2002 for dicussion). Estimates of broad-sense heritability (H^2) and their 95% confidence intervals, including all clonal replicates, was calculated as

$$H^2 = V_G / (V_G + V_E)$$

where V_G and V_E are genetic and environmental variance components, using the heritability function in the R package 'Heritability'. To estimate population differentiation, QST , the following formula was used:

$$Q_{ST}=V_{pop}/(V_{pop}+(2*V_{geno}))$$

where V_{pop} is the population and V_{geno} is the genotype genetic variance components.

Genetic correlations between phenotypes were calculated as

$$r_{G(AB)} = V_{G(AB)}/ \sqrt{(V_{G(A)} \times V_{G(B)})}$$

where $r_{G(AB)}$, the genetic correlation of phenotype A and phenotype B, was calculated from the $V_{G(AB)}$, the genetic covariance in phenotype A and phenotype B, $V_{G(A)}$ and $V_{G(B)}$ were the genetic variances of phenotypes A and B respectively.

Genetic (clonal) variation for each phenotype between years and common gardens were investigated using separate analyses of variance (ANOVA) models where phenotype was the dependent variable. Analyses were conducted only on genotypes with three or more replicate trees per genotype. To examine common garden effects in the same year, garden, genotype, and their interaction were considered independent variables in the following models:

$$Phenotype_{2008} \sim Garden + Genotype + Garden \times Genotype$$

$$Phenotype_{2011} \sim Garden + Genotype + Garden \times Genotype$$

where $Phenotype_{Year}$ indicates that the analysis was conducted on the phenotypic data from one year to compare the two gardens. To compare the phenotypic data from the two gardens in one year, the phenotypic response for a given garden (Ekebo or Sävar) was partitioned into variance of the independent variables year and genotype and their interaction:

$$Phenotype_{Ekebo} \sim Year + Genotype + Year \times Genotype$$

$$Phenotype_{Sävar} \sim Year + Genotype + Year \times Genotype$$

where $Phenotype_{Garden}$ indicates phenotypic data were taken from only one garden in each model. ANOVA models were implemented in the aov function in R. All effects were considered significant at $P<0.05$.

RNA-Sequencing data

The RNA-Seq data used in this study has been described previously (Paper III). It consists of 219 samples distributed among 86 distinct genotypes. The same type of gene expression filtering and adjustment were used in this paper as in Paper III. Genes were required to have an expression variance above 0.05, and the first nine gene expression principal components were regressed out from the data. This left 22,306 genes for further analysis. The data has been uploaded to the European Nucleotide Archive (ENA) with accession number ERP014886.

Genome wide association mapping

A total of 4.5 million SNPs were considered for the GWA, previously described in Wang *et al.* (2016, in prep). A univariate linear mixed model was applied to the data using GEMMA [35] and included the first principal component based on independent SNPs as a random effect in order to account for population structure (Wang *et al.* 2016 in prep) as well as the built-in estimation of a relatedness matrix. GEMMA produces different statistics for significance, and in this study we used p-values based on a likelihood ratio test. These p-values were consequently Benjamini-Hochberg adjusted for multiple testing for each garden and year separately using the p.adjust function in R. To associate genes with SNPs, the v1.0 *Populus tremula* annotations from the PopGenIE.org web resource were used [36], and any gene within 2 kbp of a SNP were said to be associated with that SNP.

Gene set enrichment analysis

Gene set enrichment analysis was implemented in R according to [29]. In short, a gene set was tested for significant association to a phenotype based on gene expression correlation. The expression profile of each of the 22,306 genes were correlated to the phenotype and sorted by correlation (from positive to negative values). A running sum was produced from the top of the list by adding the correlation to the sum if the gene is part of the gene set and subtracting if it is not. The test statistic is then the maximum absolute value of this running sum. This value is also used to represent the leading edge of the gene set, i.e. the genes that contribute the most to the enrichment. "Geneset %" and "Total %" in Tables 1-3 represents the portion of the data that contribute the most to the enrichment (the leading edge subset). A larger value of "Geneset %" indicates that a large portion of the genes in the gene set contribute to the enrichment, and a small value of "Total %" indicates that the leading-edge subset is tightly clustered at one extreme of the correlation distribution. Significance was determined by a permutation strategy where the phenotype sample labels were permuted. This process was repeated 1000 times, and the fraction of permuted tests that had a higher score than the score from the original data (one-sided test) was used as the p-value for the enrichment. The p-values for GSEA based on GO gene sets were not adjusted for multiple tests due to the high level of dependence in the data. Therefore these p-values should be interpreted with care, but can be used as a relative ranking metric of the different gene sets.

Gene Ontology enrichment

The R-package topGO (http://bioconductor.org/packages/release/bioc/html/topGO.html) was used to perform GO enrichment analysis. In all cases the background used for the enrichments were the set of expressed genes (22,306 genes), and the classic test was used with the Fisher test statistic. In order for a GO term to be considered enriched, the gene set tested had to contain at least two genes annotated to that particular term, regardless of p-value.

Acknowledgements

NRS is supported by the Trees and Crops for the Future (TC4F) project. The authors also would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI), and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure.

References

1. Carroll SB: **Endless forms: the evolution of gene regulation and morphological diversity.** *Cell* 2000, **101**:577–580.

2. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast**. *Science (80-)* 2002, **296**:752–755.

3. Ayroles JF, Carbone MAA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC: **Systems genetics of complex traits in Drosophila melanogaster.** *Nat Genet* 2009, **41**:299–307.

4. Mackay TFC, Stone EA, Ayroles JF: **The genetics of quantitative traits: challenges and prospects**. *Nat Rev Genet* 2009, **10**:565–577.

5. Liao B-Y, Weng M-P, Zhang J: **Contrasting genetic paths to morphological and physiological evolution**. *Proc Natl Acad Sci* 2010, **107**:7353–7358.

6. Hines HM, Papa R, Ruiz M, Papanicolaou A, Wang C, Nijhout HF, McMillan WO, Reed RD: **Transcriptome analysis reveals novel patterning and pigmentation genes underlying Heliconius butterfly wing pattern variation**. *BMC Genomics* 2012, **13**:288.

7. Richards CLC, Rosas U, Banta J, Bhambhra N, Purugganan MMD, Gibson G, Pigliucci M, Schlichting C, Jones C, Schwenk K, Leakey A, Ainsworth E, Bernard S, Markelz C, Ort D, Kammenga J, Herman M, Ouborg N, Johnson L, Breitling R, Chapman M, Leebens-Mack J, Burke J, Richards CLC, Hanzawa Y, Ehrenreich I, Purugganan MMD, Anderson JJ, Mitchell-Olds T, Boer T De, et al.: **Genome-Wide Patterns of Arabidopsis Gene Expression in Nature**. *PLoS Genet* 2012, **8**:e1002662.

8. Wang RL, Stec A, Hey J, Lukens L, Doebley J: **The limits of selection during maize domestication.** *Nature* 1999, **398**(March):236–239.

9. Civelek M, Lusis AJ: **Systems genetics approaches to understand complex traits.** *Nat Rev Genet* 2013, **15**:34–48.

10. Swami M: **Systems genetics: Networking complex traits**. *Nat Rev Genet* 2009, **10**:219–219.

11. Hardy J, Singleton A: **Genomewide Association Studies and Human Disease**. *N Engl J Med* 2009, **360**:1759–1768.

12. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: **Finding the missing heritability of complex diseases.** *Nature* 2009, **461**:747–753.

13. Goldstein DB: **Common Genetic Variation and Human Traits**. *N Engl J Med* 2009, **360**:1696–1698.

14. Cirulli ET, Goldstein DB: **Uncovering the roles of rare variants in common disease through whole-genome sequencing.** *Nat Rev Genet* 2010, **11**:415–425.

15. Ingvarsson PK, Street NR: **Association genetics of complex traits in plants.** *New Phytol* 2011, **189**:909–22.

16. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KCC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WOC: **A genome-wide association study of global gene expression**. *Nat Genet* 2007, **39**:1202–1207.

17. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET: **Population genomics of human gene expression.** *Nat Genet* 2007, **39**:1217–1224.

18. Veyrieras J-BB, Kudaravalli S, Kim SYY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK: **High-resolution mapping of expression-QTLs yields insight into human gene regulation.** *PLoS Genet* 2008, **4**.

19. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET: **Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations**. *PLoS Genet* 2010, **6**.

20. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: **Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS**. *PLoS Genet* 2010, **6**.

21. Ingvarsson PK: **Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (Populus tremula L., Salicaceae)**. *Genetics* 2005, **169**:945–953.

22. Wang J, Street NR, Scofield DG, Ingvarsson PK: **Natural Selection and Recombination Rate Variation Shape Nucleotide Polymorphism Across the Genomes of Three Related Populus Species.** *Genetics* 2015, **202**:1185–1200.

23. Wang J, Street NR, Scofield DG, Ingvarsson PK: **Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens**. *Mol Biol Evol* 2016:msw051.

24. Hall D, Luquez V, Garcia VMM, St Onge KRR, Jansson S, Ingvarsson PKK: **ADAPTIVE POPULATION DIFFERENTIATION IN PHENOLOGY ACROSS A LATITUDINAL GRADIENT IN EUROPEAN ASPEN (POPULUS TREMULA, L.): A COMPARISON OF NEUTRAL MARKERS, CANDIDATE GENES AND PHENOTYPIC TRAITS.** *Evol Int J Org Evol* 2007.

25. Ingvarsson PKK, Garcia V, Luquez V, Hall D, Jansson S: **Nucleotide polymorphism and**

phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). *Genetics* 2008, **178**:2217–2226.

26. Bylesjö M, Segura V, Soolanayakanahally RY, Rae AM, Trygg J, Gustafsson P, Jansson S, Street NR: **LAMINA: a tool for rapid quantification of leaf size and shape parameters.** *BMC Plant Biol* 2008, **8**:82.

27. Peppe DJ, Royer DL, Cariglino B, Oliver SY, Newman S, Leight E, Enikolopov G, Fernandez-Burgos M, Herrera F, Adams JM, Correa E, Currano ED, Erickson JM, Hinojosa LF, Hoganson JW, Iglesias A, Jaramillo CA, Johnson KR, Jordan GJ, Kraft NJB, Lovelock EC, Lusk CH, Niinemets U, Peñuelas J, Rapson G, Wing SL, Wright IJ, Niinemets Ü, Peñuelas J, Rapson G, et al.: **Sensitivity of leaf size and shape to climate: global patterns and paleoclimatic applications**. *New Phytol* 2011, **190**:724–739.

28. Luquez V, Hall D, Albrectsen BR, Karlsson J, Ingvarsson P, Jansson S: **Natural phenological variation in aspen (Populus tremula): the SwAsp collection**. *Tree Genet Genomes* 2008, **4**:279–292.

29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545–15550.

30. Chitwood DH, Sinha NR: **Evolutionary and Environmental Forces Sculpting Leaf Development**. *Curr Biol* 2016, **26**:R297–R306.

31. Kidner CA, Umbreen S: **Why is Leaf Shape so Variable?** *Int J Plant Dev Biol* 2010, **4**:64–75.

32. Robinson K, Ingvarsson P, Jansson S, Albrectsen B: **Genetic Variation in Functional Traits Influences Arthropod Community Composition in Aspen (Populus tremula L.)**. *PLoS One* 2012, **7**:e37679.

33. **R: A language and environment for statistical computing** [http://www.r-project.org]

34. Dohm RM: **Repeatability estimates do not always set an upper limit to heritability**. *Funct Ecol* 2002, **16**:273–280.

35. Zhou X, Stephens M: **Efficient multivariate linear mixed model algorithms for genome-wide association studies.** *Nat Methods* 2014, **advance on**.

36. Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin Y-C, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street NR: **The Plant Genome Integrative Explorer Resource: PlantGenIE.org**. *New Phytol* 2015, **208**:1149–1156.

Figures and Tables

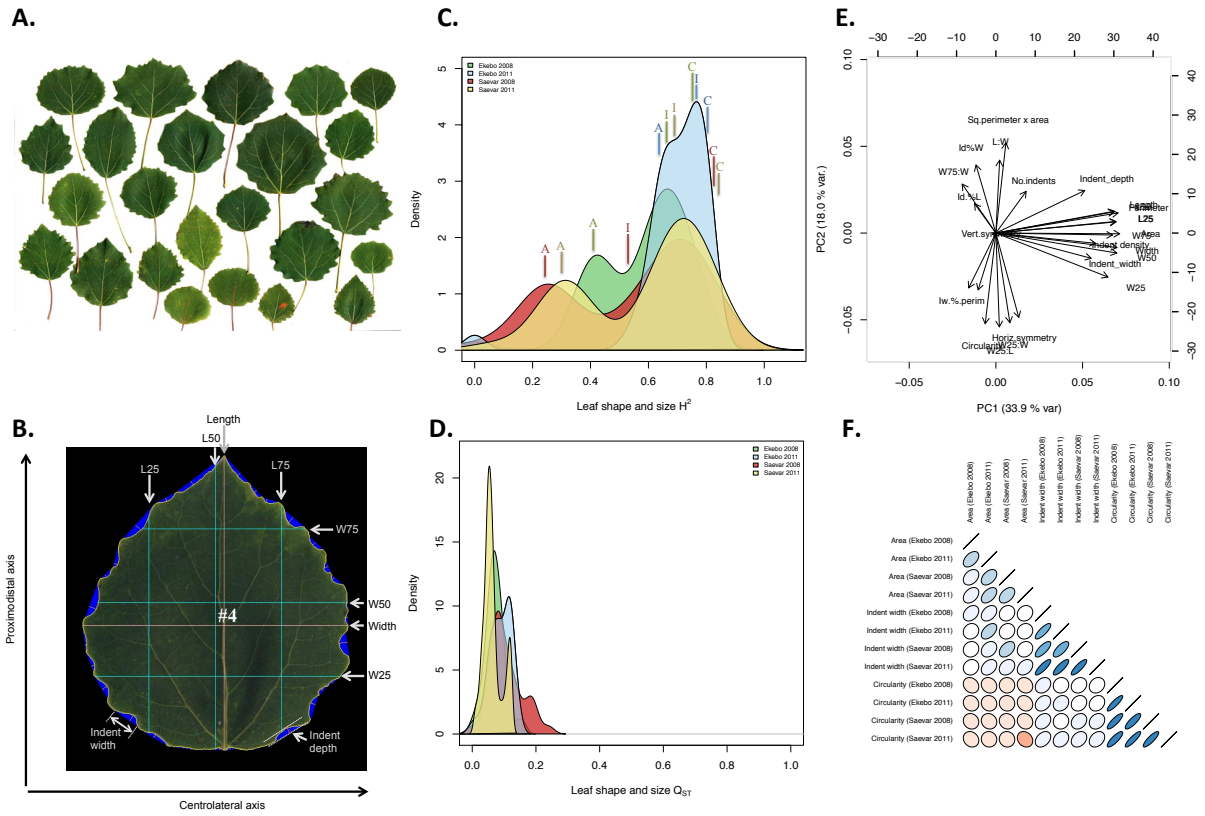

Figure 1 Leaf physiognomy variation in the Swedish Aspen collection. (A.) Natural genetic variation of physiognomic traits is apparent in a selection of the SwAsp genotypes. (B.) Leaf size metrics as measured by LAMINA (Bylesjö et al, 2008). Density distribution of (C.) heritability (H2) and (D.) QST values, with independent distributions for Ekebo 2008 (green), Ekebo 2011 (blue), Sävar 2008 (red) and Sävar 2011 (gold). Arrows on the H2 distributions indicate the H2 values for the phenotypic traits. A = leaf area, I = indent width, and C = circularity. Biplot (E.) of the first two components of principal component analysis of all size and shape phenotypes in all years and gardens. Correlation plot (E) showing the positive (blue) or negative (red) correlations between selected phenotypes in all years and gardens. Narrower ellipses represented higher correlation r values.

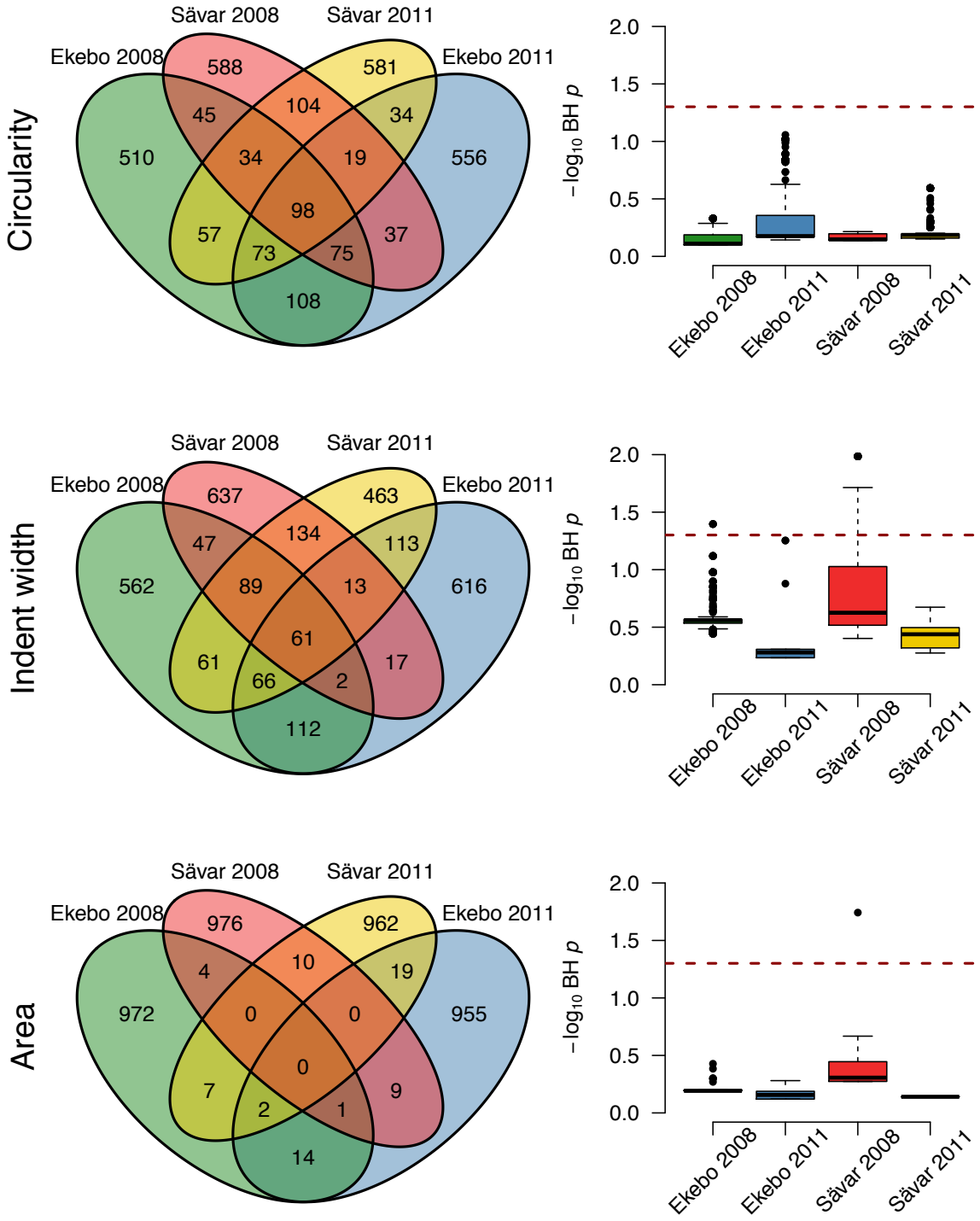

Figure 2. GWAS results for the three traits in the different gardens and years. The Venn diagrams show the top 1000 pSNPs based on p-value for each garden and year and how they overlap. The boxplots show the distribution of adjusted p-values for pSNPs that are found among the top 1000 SNPs in at least two of the gardens/years. The dashed line represents 5% FDR.

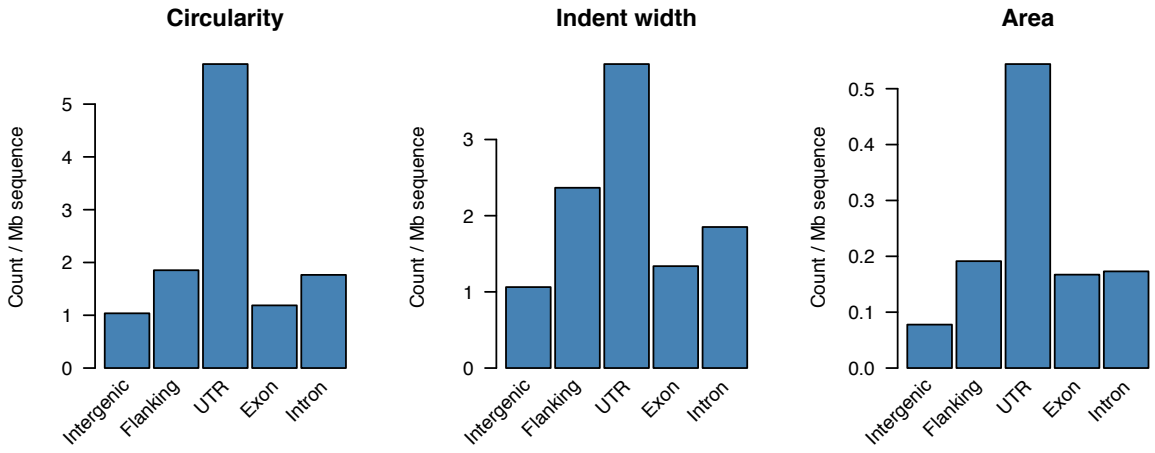

Figure 3. Genomic context of pSNPs found in at least two gardens/years (Figure 2) normalised by total feature length. In cases where SNPs overlapped several features, they were prioritised in the following way: UTR, exon, intron, flanking, intergenic. Exon counts should thus represent coding SNPs.

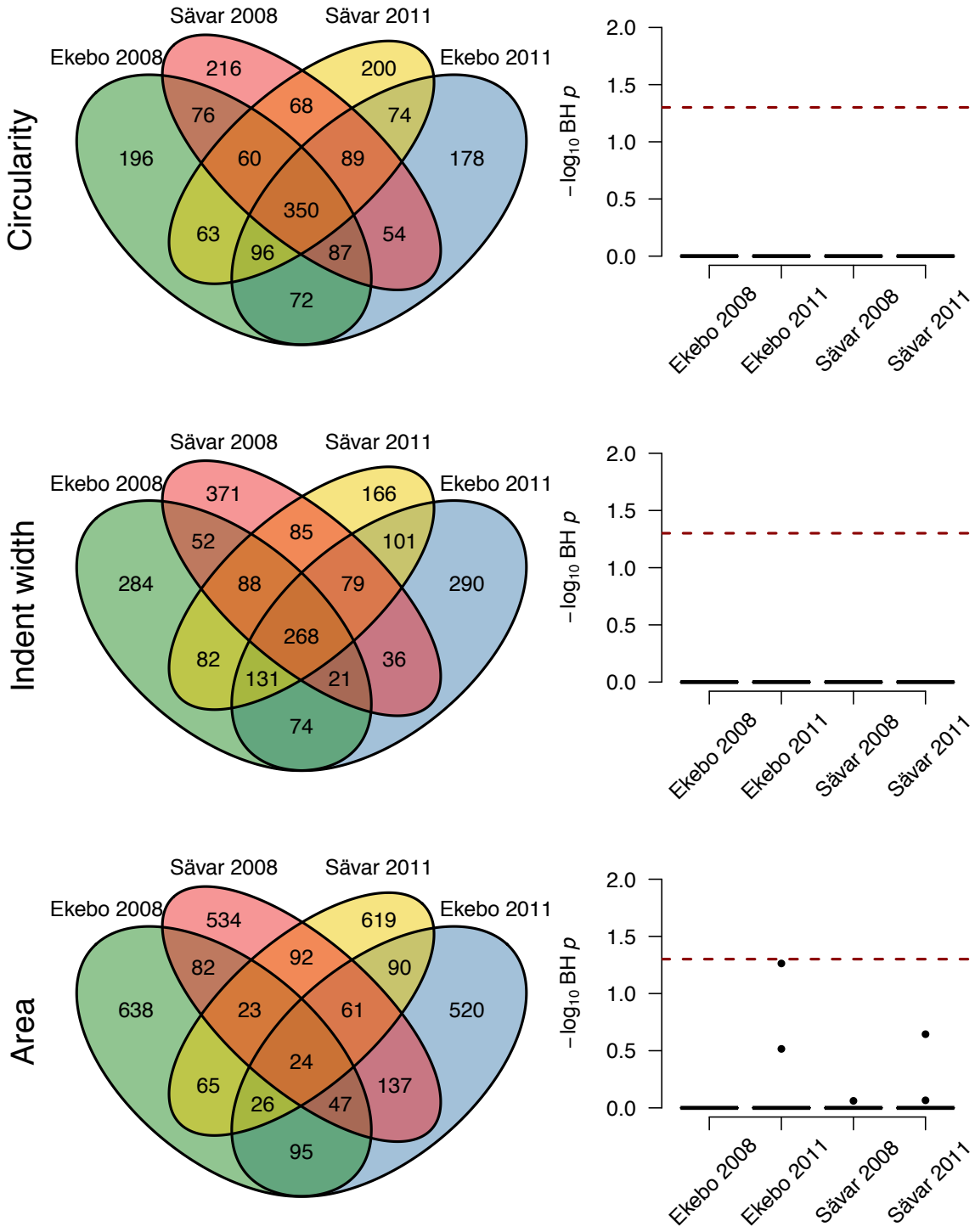

Figure 4. Correlation between gene expression profiles and the three traits for the different gardens and years. The Venn diagrams show the top 1000 epGenes based on Pearson correlation for each garden and year and how they overlap. The boxplots show the distribution of Benjamini-Hochberg adjusted p-values for correlations of the genes that are found among the top 1000 genes in at least two of the gardens/years. The dashed line represents 5% FDR.

Table 1. GSEA results for circularity. Included are GO terms with a p-value < 0.05 in at least two of the populations. The statistics presented here is the enrichment with the lowest p-value. Included are also the gene sets based on genes located within 2kb of the pSNPs discovered in at least two of the gardens/years ("pairs") and all four gardens/years ("intersection") (Figure 4). The columns are Name: name of the gene set; Description: description of the gene set; Size: the number of genes in the geneset that were expressed (Materials and methods); # eGenes: number of eGenes in the gene set; # pSNP genes: the number of genes in the gene set that were located within 2kb of a pSNP; # pSNPs: the number of pSNPs associated with the genes in the gene set; # eSNPs: the number of pSNPs that were also eSNPs; p-value: p-value of the enrichment; Geneset %: percent of the gene set that is part of the leading edge subset; Total %: the percentage of all genes that are included in the leading edge subset.

Name	Description	Size	# eGenes	# pSNP genes	# pSNPs	# eSNPs	p-value	Geneset %	Total %
GO:0008654	phospholipid biosynthetic process	11	4	0	0	0	0.001	45.45	13.43
GO:0003333	amino acid transmembrane transport	20	7	0	0	0	0.002	55.00	20.13
GO:0004722	protein serine/threonine phosphatase activity	35	8	1	2	0	0.002	54.29	20.86
GO:0006470	protein dephosphorylation	53	11	1	2	0	0.003	50.94	27.71
GO:0016829	lyase activity	25	9	0	0	0	0.003	36.00	13.48
GO:0010333	terpene synthase activity	19	7	0	0	0	0.004	36.84	9.26
GO:0006281	DNA repair	66	21	0	0	0	0.004	31.82	15.09
GO:0016310	phosphorylation	15	2	0	0	0	0.006	40.00	11.45
GO:0007165	signal transduction	197	52	2	11	0	0.007	35.53	25.02
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	16	2	0	0	0	0.009	31.25	11.58
GO:0015171	amino acid transmembrane transporter activity	15	3	0	0	0	0.01	66.67	20.13
GO:0003779	actin binding	28	9	0	0	0	0.012	42.86	23.07
GO:0030246	carbohydrate binding	106	28	0	0	0	0.012	39.62	24.33
GO:0005515	protein binding	2774	724	31	113	0	0.015	26.06	22.51
GO:0016301	kinase activity	32	12	0	0	0	0.021	31.25	10.49
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	32	14	0	0	0	0.022	53.13	30.50
GWAS SNP-gene associations	intersection	33	12	33	135	0	0.027	39.39	22.35
GO:0008408	3'-5' exonuclease activity	15	5	0	0	0	0.03	46.67	16.70
GO:0036459	ubiquitinyl hydrolase activity	21	5	0	0	0	0.03	42.86	24.01
GO:0016746	transferase activity, transferring acyl groups	36	12	0	0	0	0.043	50.00	23.56
GO:0050661	NADP binding	53	14	0	0	0	0.043	39.62	21.99
GWAS SNP-gene associations	pairs	195	54	195	484	2	0.098	26.15	20.50

Table 2. GSEA results for indent width. See Table 1 for column name descriptions.

Name	Description	Size	# eGenes	# pSNP genes	# pSNPs	# eSNPs	p-value	Geneset %	Total %
GO:0015171	amino acid transmembrane transporter activity	15	3	0	0	0	0.001	80.00	29.99
GO:0006812	cation transport	49	17	0	8	0	0.001	38.78	22.69
GO:0003779	actin binding	28	9	0	0	0	0.002	53.57	17.40
GO:0016762	xyloglucan:xyloglucosyl transferase activity	23	4	1	1	0	0.003	47.83	10.98
GO:0003333	amino acid transmembrane transport	20	7	0	0	0	0.004	70.00	30.85
GO:0004842	ubiquitin-protein transferase activity	84	12	2	2	0	0.004	34.52	17.90
GO:0006855	drug transmembrane transport	47	12	0	0	0	0.005	36.17	14.58
GO:0008233	peptidase activity	18	4	0	0	0	0.006	44.44	22.89
GO:0016746	transferase activity, transferring acyl groups	36	12	0	0	0	0.007	41.67	19.65
GO:0004222	metalloendopeptidase activity	32	11	0	0	0	0.008	46.88	23.21
GO:0006073	cellular glucan metabolic process	23	4	1	1	0	0.01	47.83	10.98
GO:0048544	recognition of pollen	74	26	0	0	0	0.011	36.49	22.01
GO:0001522	pseudouridine synthesis	15	6	0	0	0	0.021	73.33	28.53
GO:0005874	microtubule	17	5	0	0	0	0.024	52.94	12.92
GO:0005488	binding	311	82	1	1	0	0.025	33.76	23.38
GO:0005200	structural constituent of cytoskeleton	17	5	0	0	0	0.026	47.06	7.46
GO:0048046	apoplast	49	6	1	1	0	0.028	40.82	23.14
GO:0008272	sulfate transport	15	1	1	5	0	0.029	26.67	11.76
GO:0007165	signal transduction	197	52	1	5	0	0.042	27.92	20.51
GWAS SNP-gene associations	pairs	160	40	160	510	31	0.086	16.25	10.23
GWAS SNP-gene associations	intersection	13	1	13	112	4	0.135	23.08	9.55

Table 3. GSEA results for area. See Table 1 for column name descriptions.

Name	Description	Size	# eGenes	# pSNP genes	# pSNPs	# eSNPs	p-value	Geneset %	Total %
GO:0016762	xyloglucan:xyloglucosyl transferase activity	23	4	0	0	0	0.001	60.87	21.65
GO:0006139	nucleobase-containing compound metabolic process	34	16	0	0	0	0.002	50.00	22.98
GO:0030145	manganese ion binding	24	7	0	0	0	0.002	50.00	21.51
GO:0000413	protein peptidyl-prolyl isomerization	33	12	1	1	0	0.002	48.48	25.96
GO:0005488	binding	311	82	0	0	0	0.002	35.37	23.24
GO:0006073	cellular glucan metabolic process	23	4	0	0	0	0.003	47.83	13.11
GO:0006259	DNA metabolic process	14	6	0	0	0	0.004	64.29	28.89
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	33	12	1	1	0	0.004	45.45	20.40
GO:0006629	lipid metabolic process	130	34	0	0	0	0.009	43.08	27.80
GO:0008408	3'-5' exonuclease activity	15	5	0	0	0	0.01	60.00	18.24
GO:0000287	magnesium ion binding	90	31	1	3	0	0.012	45.56	24.73
GO:0006006	glucose metabolic process	16	6	0	0	0	0.012	43.75	20.17
GO:0006281	DNA repair	66	21	0	0	0	0.012	28.79	17.57
GO:0016310	phosphorylation	15	2	0	0	0	0.02	46.67	24.51
GO:0048046	apoplast	49	6	0	0	0	0.021	26.53	9.36
GO:0022891	substrate-specific transmembrane transporter activity	49	13	0	0	0	0.034	34.69	16.96
GO:0008233	peptidase activity	18	4	0	0	0	0.041	50.00	22.08
GWAS SNP-gene associations	pairs	28	5	28	51	0	0.052	28.57	15.61

Supplementary file S1: Microsoft Excel file with six sheets. Available at https://figshare.com/s/c40f29949c545f788e22.

Sheet 1. **H2.** Broad-sense heritability (H^2) values, and upper and lower 95% confidence intervals for all leaf shape and size phenotypes and their PCA summaries.

Sheet 2. **QST.** Q_{ST} values, and upper and lower 95% confidence intervals for all leaf shape and size phenotypes and their PCA summaries.

Sheet 3. **Ekebo 2008 x 2011.** Analyses of variance tables for all leaf shape and size phenotypes, from the model: $Phenotype_{Ekebo} \sim Year + Genotype + Year \times Genotype$. Analyses were conducted only on genotypes with three or more replicate trees per genotype.

Sheet 4. **Saevar 2008 x 2011.** Analyses of variance tables for all leaf shape and size phenotypes, from the model: $Phenotype_{Saevar} \sim Year + Genotype + Year \times Genotype$. Analyses were conducted only on genotypes with three or more replicate trees per genotype.

Sheet 5. **2008 Saevar x Ekebo.** Analyses of variance tables for all leaf shape and size phenotypes, from the model: $Phenotype_{2008} \sim Garden + Genotype + Garden \times Genotype$. Analyses were conducted only on genotypes with three or more replicate trees per genotype.

Sheet 6. **2011 Saevar x Ekebo.** Analyses of variance tables for all leaf shape and size phenotypes, from the model: $Phenotype_{2011} \sim Garden + Genotype + Garden \times Genotype$. Analyses were conducted only on genotypes with three or more replicate trees per genotype.

Supplementary file S2: Microsoft Excel file with three sheets. Available at https://figshare.com/s/d952ccec6a3a26f60795.

Sheet 1. **Leaf shape and size metrics.** Leaf metrics for leaf shape and size phenotypes are listed separately, with their definitions. For further details, please see [26].

Sheet 2. **Sampling information.** A brief summary of the numbers of trees and the numbers of genotypes represented in each common garden and sampling year.

Sheet 3. **PCA loadings.** Principal component analyses loadings for the first components of **(a)** shape, **(b)** size and **(c)** shape and size phenotypes.