Genomic regions associate with major axes of variation driven by gas exchange and leaf construction traits in cultivated sunflower (Helianthus annuus L.)

Running title: Genomics of leaf trait variation in sunflower

Ashley M. Earley[1], Andries A. Temme[1,2], Christopher R. Cotter[1], John M. Burke[1]

[1] *Department of Plant Biology, University of Georgia, Athens, GA, USA*

[2] *Division of Intensive Plant Food Systems, Humboldt-Universität zu Berlin, 10117 Berlin, Germany*

Abstract

Stomata and leaf veins play an essential role in transpiration and the movement of water throughout leaves. These traits are thus thought to play a key role in the adaptation of plants to drought and a better understanding of the genetic basis of their variation and coordination could inform efforts to improve drought tolerance. Here, we explore patterns of variation and covariation in leaf anatomical traits and analyze their genetic architecture via genome-wide association (GWA) analyses in cultivated sunflower (*Helianthus annuus* L.*)*. Traits related to stomatal density and morphology as well as lower order veins were manually measured from digital images while the density of minor veins was estimated using a novel deep learning approach. Leaf, stomatal, and vein traits exhibited numerous significant correlations that generally followed expectations based on functional relationships. Correlated suites of traits could further be separated along three major principal component (PC) axes that were heavily influenced by variation in traits related to gas exchange, leaf hydraulics, and leaf construction. While there was limited evidence of colocalization when individual traits were subjected to GWA analyses, major multivariate PC axes that were most strongly influenced by several traits related to gas exchange or leaf construction did exhibit significant genomic associations. These results provide insight into the genetic basis of leaf trait covariation and showcase potential targets for future efforts aimed at modifying leaf anatomical traits in sunflower.

Significance Statement:

Using traditional and automated/high-throughput (using a novel deep learning approach) phenotyping methods we studied leaf anatomical variation in sunflower. Genome-wide association (GWA) analyses identified numerous genomic regions underlying individual trait variation and regions underlying major multivariate axes of phenotypic variation. These results illustrate the value of employing a multivariate approach to GWA analyses and shed light on the extent to which leaf trait (co-)variation can be genetically decoupled to explore novel phenotypic space.

Keywords: leaf anatomy, stomata, venation, GWAS, sunflower, deep learning, neural network

Introduction

Stomata and leaf veins are central players in plant water relations. Veins distribute water throughout the leaf and stomata control the rate of transpiration (Sack and Scoffoni, 2013). As such, these traits are generally thought to play a key role in the adaptation of plants to drought (e.g., Bertolino *et al.*, 2019; Sack and Scoffoni, 2013), a major agricultural stress that limits plant growth and productivity worldwide (NOAA.; IPCC, 2014). Plant performance in water-limited conditions is known to be influenced by stomatal and vein traits (Lei *et al.*, 2018; Scoffoni *et al.*, 2011; Buckley, 2019). Stomatal conductance (g_s), a measure of the rate of CO_2 uptake and water loss through the stomata, is determined by the physiological control of stomatal opening and closing as well as the size and distribution of stomata (Faralli *et al.*, 2019). Similarly, vein length per area (VLA) affects leaf hydraulic conductance (K_{leaf}), or the resistance of water flow through the leaf, and thus greatly affects how long stomata can remain open without drying out the leaf (Sack and Scoffoni, 2013). While substantial work has been done on the developmental and regulatory pathways determining stomatal density and morphology (e.g., Gudesblat *et al.*, 2012; Bergmann and Sack, 2007) less is known about the genetic basis of the observed relationships between stomatal and leaf vein traits.

Leaf veins distribute water from the petiole throughout the leaf. In the intracellular space, water then moves through the mesophyll to the ultimate site of transpiration, the stomata. It is important to note here that stomata can be distributed on both the top (adaxial) and bottom (abaxial) sides of the leaf, and that species with stomata on both sides (i.e., amphistomatous) tend to have greater gas exchange capacity than species with stomata on a single side (Xiong and Flexas, 2020). An increase in the number of stomata on the upper side leads to increased maximum photosynthetic rates and to increased rates of transpiration due to increased CO_2 diffusion (Muir, 2018; Xiong and Flexas, 2020). Due to these effects on water supply and loss, a close relationship between vein density and stomatal size/density is expected to be important for the optimization of water transport and transpiration (Fiorin *et al.*, 2016; Bertolino *et al.*, 2019). In particular, a shorter vein-to-stomata distance is thought to improve gas exchange and photosynthetic performance (Brodribb *et al.*, 2007; de Boer *et al.*, 2012). This distance is affected by traits such as stomatal density, vein density (estimated as vein length per area; VLA), leaf thickness, and cell shape (Brodribb *et al.*, 2007; de Boer *et al.*, 2012). As a general rule, vein and stomatal density are positively correlated (Sack and Scoffoni, 2013) while stomatal size and density are negatively correlated (Shahinnia *et al.*, 2016; Doheny-Adams *et al.*, 2012). Such correlations are observed within (Carins Murphy *et al.*, 2014) and across species (Zhang *et al.*, 2012), though the extent to which these trait correlations are conditioned by genetic correlations (i.e., linkage or pleiotropy of major effect loci) remains an open question.

At the whole leaf level, plant species exhibit a range of strategies related to the cost of leaf construction. Traits such as leaf mass per area (LMA) and VLA play a role in leaf construction in addition to affecting leaf hydraulics (Xing *et al.*, 2021; Poorter *et al.*, 2009; John *et al.*, 2017). These strategies occur along a major axis of leaf trait variation (the Leaf Economics Spectrum, LES), which ranges from resource conservative (i.e., ‘slow’) with a greater investment in leaf construction to resource acquisitive (i.e., ‘fast’) with a smaller investment in leaf construction (Wright *et al.*, 2004; Reich, 2014; Díaz *et al.*, 2016). A key indicator trait for the location of

species along this spectrum is LMA, with leaf hydraulic traits (allocation to major vs. minor veins in particular) accounting for a portion of the observed variation in LMA (John *et al.*, 2017). Investigating the variation present in both finer scale aspects of leaves as well as whole leaf traits promises to broaden our understanding of the evolution of correlated suites of leaf traits.

To date, most studies focusing on within species variation in leaf anatomy have either relied on relatively small sample sizes or have largely focused on stomatal traits to the exclusion of vein traits (e.g., Khan *et al.*, 2003; Ries *et al.*, 2012; Shi *et al.*, 2021; Haworth *et al.*, 2021). The general dearth of studies that explicitly examine the genetic basis of stomatal and vein traits, particularly in tandem, is likely due to challenges associated with the large-scale phenotyping of such traits. One study on wild tomato found that strongly correlated leaf traits were not controlled by the same QTL, suggesting that natural selection had favored particular trait combinations (Muir *et al.*, 2014). However, another study on wild *Arabidopsis* accessions found that stomatal density correlated with various other leaf traits, including stomatal index and pavement cell density, and that they seemed to share a common genetic basis (Delgado *et al.*, 2011). The lack of a clear pattern across species leaves as an open question the extent to which variation in such traits can be decoupled, thereby allowing them to vary independently.

Here, we describe patterns of phenotypic variation in leaf traits in cultivated sunflower (*Helianthus annuus* L.) and investigate their genetic architecture via genome-wide association (GWA) analyses. Domesticated from the common sunflower (also *H. annuus*; Wills and Burke, 2006; Blackman *et al.*, 2011), cultivated sunflower is one of the world's most important oilseed crops (FAO, 2018). Often grown in rainfed regions, sunflower productivity is frequently dependent on natural patterns of precipitation. While sunflower is generally regarded as being drought resistant due to its ability to root deeply (Connor and Sadras, 1992), drought is considered to be a major yield-limiting factor across the range of production (Hussain *et al.*, 2018), making traits underlying plant-water relations a vital avenue of research. Prior work on sunflower has shown substantial variability and plasticity in leaf anatomical traits (Wang *et al.*, 2020). Previous genomic analyses however, have been largely limited to higher level traits such as leaf mass, leaf area, LMA, and leaf mass fraction (Temme *et al.*, 2020; Masalia *et al.*, 2018). To overcome limitations in the analysis of vein traits, we developed a neural network deep learning approach to increase phenotyping efficiency from digital images (see Xu *et al.*, 2020 for a similar approach), enabling an investigation of the genomic basis on finer scale leaf anatomical traits.

In this study, we sought to: (1) quantify phenotypic variation in stomatal and leaf vein traits and test for trait correlations in sunflower, (2) identify genomic regions underlying these traits using genome-wide association analyses, and (3) determine the extent to which observed trait correlations are due to a shared genetic basis. Our results provide insight into the genetic complexity of these traits and the degree to which observed trait correlations are constrained by linkage or pleiotropy of major effect loci and serve as a valuable first step toward optimizing leaf trait combinations via breeding.

Methods

Plant Material:

The cultivated sunflower lines analyzed in this study comprise the sunflower association mapping (SAM) population (Mandel *et al.*, 2011), which includes 288 inbred lines that capture ca. 90% of allelic diversity in crop sunflower (Mandel *et al.*, 2013). This population has since been subjected to whole genome re-sequencing, thereby enabling the identification of a genome-wide collection of single nucleotide polymorphisms (SNPs) from the full set of lines (Hübner *et al.*, 2019).

Experimental Design:

In the summer of 2017, 239 inbred lines from the SAM population (4 four replicates each; N = 4 x 239 = 956 total individuals) were grown in the greenhouse in a randomized block design. The plants used in this study correspond to the control plants from Temme et al. (2020) and detailed plant growth methods are described therein. Briefly, 239 of the 288 lines in the mapping population were used due to greenhouse space constraints and to remove lines with greater than expected levels of heterozygosity. Following germination, all plants were grown for one week in seedling trays to allow for establishment before being transplanted into 2.83 L pots (TP414; Stuewe & Sons, Tangent, OR) filled with a 3:1 mixture of sand and a calcined clay mixture (Turface MVP, Turface Athletics). Pots were fertilized with 40g Osmocote Plus (15-12-9 NPK; ScottsMiracle-Gro, Marysville, OH) and 5 mL each of gypsum (Performance Minerals Corporation, Birmingham, AL) and lime (Austinville Limestone, Austinville, VA) powders for supplemental Ca^{2+}. All pots were well-watered and plants were allowed to grow for three additional weeks before being harvested at four weeks old. At harvest, biomass was collected and dried in ovens at 60°C for at least 72 hours. Roots were washed and dried in the same manner. Dried samples were weighed to calculate total and aboveground biomass. During harvest, the two most recently fully expanded leaves (MRFEL) were also collected from each plant. One leaf was arbitrarily designated for use in the estimation of LMA and the other was used for leaf anatomy analyses. The adaxial and abaxial (hereafter top and bottom) surfaces of one half of one MRFEL per plant were pressed into dental putty (President Dental Putty; Coltène/Whaledent Inc., Cuyahoga Falls, OH) to create an impression of the epidermis of each leaf surface to allow for the visualization and analysis of stomatal traits following the general methods of (Weyers and Johansen, 1985). The other half of the same leaf was stored in Formalin-Acetic Acid-Alcohol (FAA) fixative for imaging and analysis of vein traits.

To estimate LMA, the designated MRFEL from each plant was scanned on a flatbed scanner at 300 dpi (Temme *et al.*, 2020). This image was then used to calculate leaf area using ImageJ v1.52b (Schindelin *et al.*, 2012). The leaf was then dried at 60°C for 48 hours and weighed (the petiole was not included). Using both the mass and area measurements, LMA was calculated as LMA = dry mass/unit area (g/m^2).

For stomatal traits, clear nail polish was applied to the epidermal impressions of the top and bottom leaf surfaces and subsequently peeled off using clear tape and placed on microscope

slides (Hilu and Randall, 1984; Weyers and Johansen, 1985). Slides were imaged using a Zeiss Axioskop 2 microscope along with ZEN software (Carl Zeiss Microscopy) under the 100X objective to enable estimation of stomatal size. A second set of images (four different fields of view per impression) were taken using the 20X objective to enable estimation of stomatal density. Size estimates were based on 10 stomata per leaf, separately for the top and bottom surfaces of each leaf, for a total of 20 stomata per plant. Stomatal length, pore length, and guard cell width were measured for each stomate using ImageJ (Figure 1A). Stomatal densities were estimated by counting the number of stomata in each of the four fields of view (counting partial stomata on only two sides of each image) per side of each leaf (eight images total per plant). Stomatal ratio was then calculated as number of bottom stomata/total stomata and stomatal sum was calculated as number of top stomata + number of bottom stomata. For consistency with the literature, we used stomatal sum instead of average density of stomates (e.g., Muir, 2018; Richardson *et al.*, 2020). Finally, maximum stomatal conductance (g_{smax}), the theoretical maximum rate of gas exchange if all stomata were fully open (calculated as sum of top and bottom), was calculated based on stomatal density and size measurements following the approach of Dow et al. (2014). This was used instead of directly measuring g_s since direct measurements were not feasible for such a large sample size.

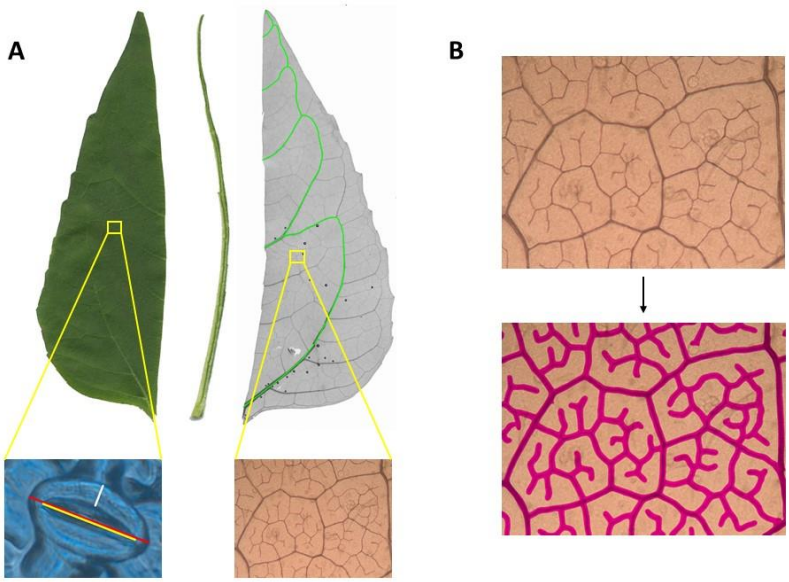

Figure 1*: (A) Left: Image of dissected leaf. Right: Cleared and stained leaf showing major veins (green). Left Inset: A single stoma taken using the100X objective. Colored lines indicate measurements taken: stomate length (red), pore length (yellow), and guard cell width (white). Right Inset: A microscope image of minor veins taken using the 5X objective. (B) Top: Image of minor veins taken at 5X. Bottom: Image of computer-traced minor veins using our deep learning approach.*

For vein traits, the half of each leaf that was fixed in FAA was cleared and stained for analysis using a modification of established procedures (Berlyn *et al.*, 1976; Scoffoni and Sack, 2013). Leaves were cleared in 5% NaOH at 55°C for 5-7 days. Subsequently, leaves were rinsed with deionized water and run through an ethanol dilution series of 30%, 50%, and 70% to dehydrate.

After dehydration, leaves were stained with a 0.01% safranin dye solution for 30 minutes to make the veins more visible. Images of the stained leaves were captured using both a flatbed scanner at 2400 dpi to image the entire leaf half and a microscope (Zeiss Axioskop 2) using the 5X objective on a small section of leaf with three to four different fields of view per leaf (Vasco *et al.*, 2014; Figure 1 A-B). Second-order vein length was measured by manually tracing veins that branched off the midrib (primary vein) and joined together in arches (Ellis *et al.*, 2009). Major vein length was estimated by adding the length of the midrib to the second-order vein length. Minor vein lengths were estimated from a subset of the microscopic images by manually tracing the veins with ImageJ. The subset that was manually traced was then used to train a deep learning algorithm (see below) to process the rest of the images. The results of these analyses of minor veins were used to calculate vein length per area (VLA). Midrib density (expressed as mg/cm^3) was estimated from the mass of the midrib and an estimate of its volume; the latter value was calculated from the length of the midrib and its diameter at the base of the leaf and assuming a conical shape. Midrib mass fraction is the ratio of mass of the midrib to mass of the leaf. A composite trait of stomata number per vein length was also calculated as SV = average stomatal density (i.e., calculated as the average of the top and bottom density estimates) divided by VLA (Zhao *et al.*, 2017).

Image Segmentation Using a Neural Network:

A modified version of the U-Net (Xu *et al.*, 2020; Ronneberger *et al.*, 2015) deep neural network was used to segment leaf-vein pixels from background pixels (Figures S1-S3). Network structure, hyperparameter tuning, and training details are included in Methods S1. Briefly, a test set of 85 images was created by randomly selecting one image from each of 85 randomly selected (without replacement) genotypes. These images were not used during training of the network or hyperparameter estimation. All remaining, manually traced (i.e., hand-segmented) images were used in the training set (747 images . Training was performed on randomly selected 572 pixels wide x 572 pixels high x 3 color (RGB) channels regions of the bright-field leaf images. Each region was normalized to a mean of 0 and a standard deviation of 1. The network was trained with a batch size of 1 (Ronneberger *et al.*, 2015) for 1900 batches with a learning rate of 10^{-4}.

Per-pixel in-vein predictions were performed by mirror padding each full size (2584x1936 pixel) bright field image, normalizing to a mean of zero and standard deviation of one, and passing the images through the trained network. To reduce the creation of small, non-real branches during medial-axis thinning, the probabilities were filtered with a Gaussian kernel with a standard deviation of 12, about the width of a typical vein. The smoothed images were segmented with a cutoff of 0.2. The resulting segmented images were thinned to single-pixel-wide lines using the medial-axis transform (Bucksch, 2014). Vein lengths were then calculated as the sum of all pixels within the thinned line. The hand-segmented and network-segmented vein lengths of the testing set images have a Pearson correlation of 0.97 (Figure S2). All code can be found at https://github.com/aatemme/burke_leaf_veins/. For additional methodological details, see Methods S1.

Data Analysis:

All data analysis was conducted using R v3.4.3 (R Core Team, 2013). Bivariate plots were made for all pairwise comparisons using the R package *ggplot2* (Wickham, 2016) to show the range of trait variation. A two-way ANOVA with genotype and block as the main effects was performed to test for variation among genotypes and to calculate estimated marginal trait means for genotypes after removing block effects. Marginal means calculated with the R package *emmeans* (Lenth, 2020) were also used to estimate trait correlations and to create a correlation matrix using the R package *corrplot (Wei and Simko, 2017)*. A principal component analysis (PCA) was conducted using the function prcomp() and the package *ggfortify* (Tang *et al.*, 2016; Horikoshi and Tang, 2018) to visualize multivariate correlations. Narrow sense heritabilities (h^2) were calculated using the package *heritability* (Kruijer *et al.*, 2019) by inputting the kinship matrix (see below) for all genotypes and individual pot-level trait data.

All traits of interest, including the first three principal components (PCs) from the PCA, were analyzed via GWA analyses to identify genomic regions that are significantly associated with each trait. These analyses were performed using a custom pipeline as described in Temme et al. (2020; https://github.com/aatemme/Sunflower-GWAS-v2). SNPs used in these analyses were called as described in Hübner et al. (2019) and reordered based on the improved HA412-HOv2 sunflower genome assembly (Todesco *et al.*, 2020). SNPs were filtered to retain those with minor allele frequency ≥ 5% and residual heterozygosity of < 10% (Temme et al., 2020). The GWA analyses were performed using GEMMA v.98.1 (Zhou and Stephens, 2012) to test for associations while correcting for kinship (as calculated by GEMMA) and population structure (using the first four PCs from an analysis using a subset of independent SNPs in SNPRelate; Zheng *et al.*, 2012). The significance threshold was corrected for multiple comparisons using a modified Bonferroni correction based on the number of multi-SNP haplotypic within the genome as determined by an analysis of linkage disequilibrium (LD) across the population (Temme *et al.*, 2020); any SNPs exceeding this threshold were considered to be significantly associated with the trait of interest. Following this step, significant SNPs were grouped into significantly associated genomic regions, with all SNPs occurring within a previously identified haplotypic block being assumed to mark a single region (see Temme et al., 2020 for details). Suggestive associations were then identified as SNPs in the top 0.1% of all SNPs that co-localized with significant associations for one or more other traits. The relative effect size (RES) of associated SNPs/regions were then estimated as the percentage of the observed range of variation in a particular trait that is explained by each association, as follows:

$$\text{RES} = |(2 * \beta/\text{range of observed trait values})| * 100$$

Here, β represents the effect of the minor allele on the trait value and the range is based on the distribution of trait values across all genotypes (Masalia *et al.*, 2018). In cases of multi-SNP blocks, the RES value was estimated as the maximum value for all SNPs within that block. These values can also be summed to represent the total percentage variation explained by all significant associations for a particular trait.

Results

Patterns of phenotypic variation and trait correlations

Significant genotypic effects were detected within the SAM population for all traits measured except midrib mass fraction (Table 1). Table 1 lists the median, minimum, and maximum values found for all traits measured, demonstrating substantial trait variation across the population. Comparing manually measured vein length per area (VLA) to the results from the neural network analysis revealed a strong correlation (Pearson's r = 0.97), demonstrating the accuracy and validity of these computer-generated measurements (Figure S2).

Table 1: *List of all traits measured along with the mean and range of trait values. Significance for genotypic effects (***P ≤ 0.001, **P ≤ 0.01, * P ≤ 0.05, ns = not significant) and adjusted R^2 from the model are also presented. SD = stomatal density; LMA = leaf mass per area; MF = mass fraction; GCW = guard cell width; SL = stomatal length; PL = stomatal pore length; g_{smax} = theoretical maximum stomatal conductance; VLA = vein length per area; SV = stomata per vein length; AG = aboveground; Bio = biomass.*

Trait	Median	Min	Max	Genotypic Effect	Adjusted R^2
SD_Bottom (stomata/mm^2)	211.20	125.17	399.17	***	0.24
SD_Top (stomata/mm^2)	158.72	86.73	288.62	***	0.21
SD_Sum (stomata/mm^2)	373.37	227.50	369.30	***	0.25
Stomatal Ratio (bottom/sum)	0.57	0.48	0.68	***	0.16
LMA (g/m^2)	22.56	12.28	32.03	***	0.33
Leaf Area (m^2)	0.0096	0.0035	0.0228	***	0.48
Midrib MF (g_{midrib}/g_{leaf})	0.07	0.03	0.10	ns	0.04
Midrib Density (mg/cm^3)	67.49	33.93	118.86	*	0.10
GCW_Bottom (µm)	6.51	5.28	8.00	**	0.09
GCW_Top (µm)	5.94	4.67	7.79	***	0.12
GCW_Avg (µm)	6.24	4.98	7.89	***	0.12
SL_Bottom (µm)	33.31	27.74	42.09	***	0.18
SL_Top (µm)	30.60	25.62	36.63	***	0.15
SL_Avg (µm)	32.09	26.81	39.36	***	0.19
PL_Bottom (µm)	24.26	19.02	31.95	***	0.21
PL_Top (µm)	21.78	16.90	29.19	***	0.16
PL_Avg (µm)	23.13	18.57	29.55	***	0.20
g_{smax} (mol/m^{2}s)	10.73	7.05	18.56	***	0.35
VLA (mm/mm^2)	9.48	7.03	12.63	***	0.23
2nd VLA (mm/mm^2)	0.07	0.02	0.14	***	0.19
Major VLA (mm/mm^2)	0.10	0.05	0.20	***	0.21
SV (stomata/mm)	19.82	12.99	31.84	***	0.24
AG Bio (g)	3.78	0.80	9.45	***	0.53
Plant Bio (g)	4.31	1.01	10.65	***	0.53

Trait correlations were examined using both bivariate and multivariate analyses. Stomatal density and size traits from the top and the bottom of each leaf were strongly correlated with each other (Figure 3A; Figure S4; e.g., SD_Top vs. SD_Bottom, PL_Top vs. PL_Bottom); as such, stomatal sum (SD_Top + SD_Bottom) and averages of the top and bottom for other traits were used for analyses moving forward. Bivariate analysis revealed numerous significant trait correlations among stomatal and vein traits (Figure 2; Figure 3). Notably, stomatal density and length were negatively correlated (Figure 3B), stomatal density and VLA were positively correlated (Figure 3C), and stomatal length and VLA were negatively correlated (Figure S4). VLA was positively correlated with both average stomatal density and theoretical g_{smax} (Figure 3D). The only traits that were significantly correlated with leaf area were second VLA and major VLA and biomass traits indicating that trait scaling with leaf size did not play a major role in observed patterns of anatomical variation. For all bivariate plots see Figure S4.

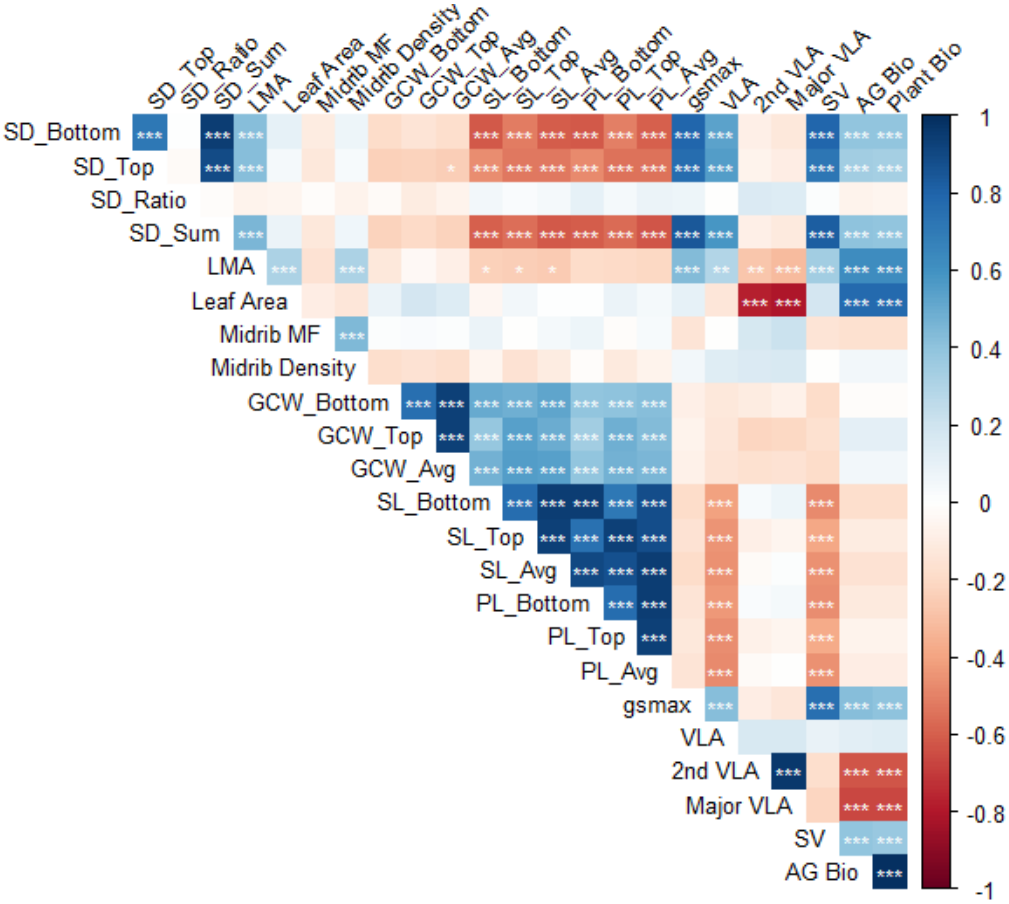

Figure 2*: Correlation matrix of leaf traits. Values were calculated using genotype marginal means and significance tests were corrected for multiple comparisons using a Bonferroni correction. Positive correlations are in blue and negative correlations are in red. Shading gives a relative indication of the magnitude of the estimate. ***P ≤ 0.001, **P ≤ 0.01, * P ≤ 0.05. Trait abbreviations are as defined in Table 1.*

Multivariate trait correlations were analyzed via PCA to determine major axes of variation. Because stomatal trait values on the top and bottom of the leaf were strongly correlated, we again used an average of the top and bottom values for stomatal size traits including length, pore length, and guard cell width (Figure 4; see Figure S5 for a full PCA including the top and bottom traits separately). The first three PCs combined to explain 62.6% of trait variation (Figure S6) with PC1 explaining 30.3%, PC2 explaining 22.6%, and PC3 explaining 10.0% of variation. The top three traits contributing to each of the major axes were: PC1 – stomatal density (SD_Sum), stomata per vein length (SV), and whole plant biomass (Plant Bio); PC2 – 2nd VLA, Major VLA, and leaf area; and PC3 – midrib density, midrib mass fraction, and LMA (Table 2). While all traits have a loading score on each PC axis, we sought to infer some functional significance for each of the major axes of variation. Given the observed trait loadings, PC1 appears to be most heavily influenced by traits related to gas exchange functioning such as stomatal density sum and stomata per vein length. In contrast, PC2 is strongly influenced by traits related to hydraulic functioning including major VLA and second VLA. Finally PC3 is heavily influenced by traits related to cost of leaf construction including LMA and midrib traits.

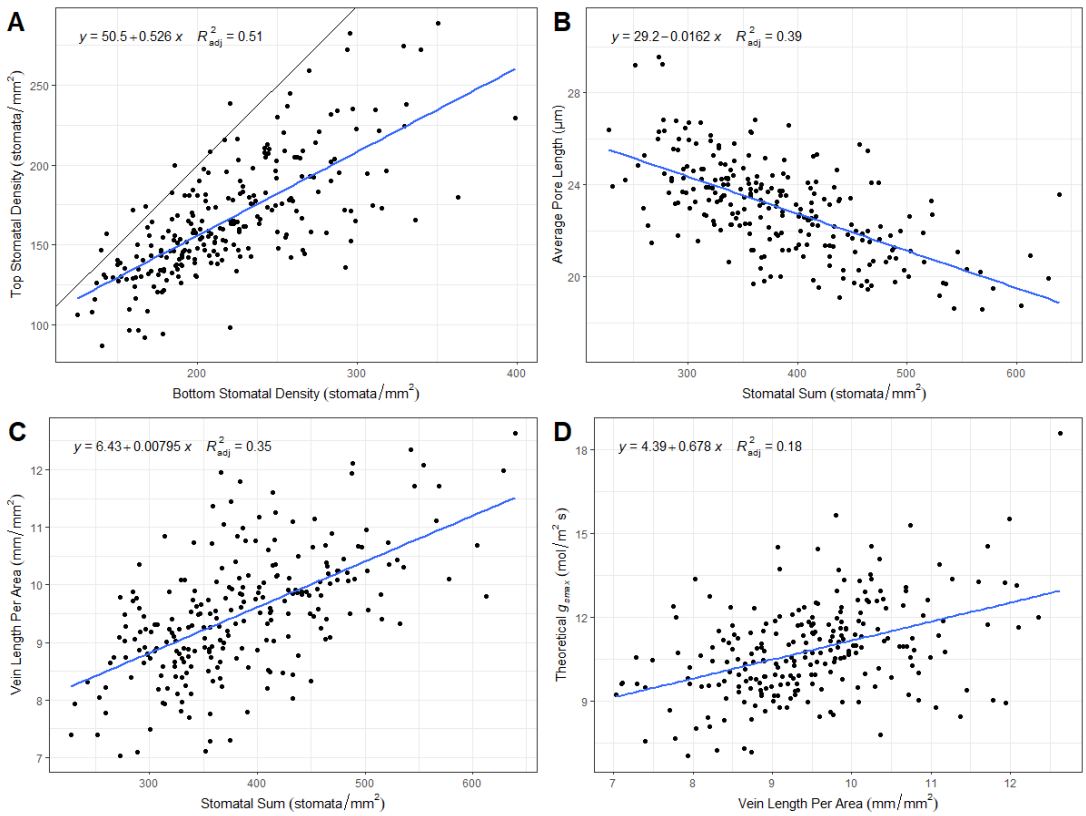

Figure 3*: Example bivariate trait plots. In all panels, the blue line is the best fit. (A) Top vs. bottom stomatal density (black line is 1:1). (B) Average (top and bottom) stomatal pore length vs. stomatal sum. (C) Vein length per area vs. stomatal sum. (D) Theoretical stomatal conductance (g_{smax}) vs. vein length per area.*

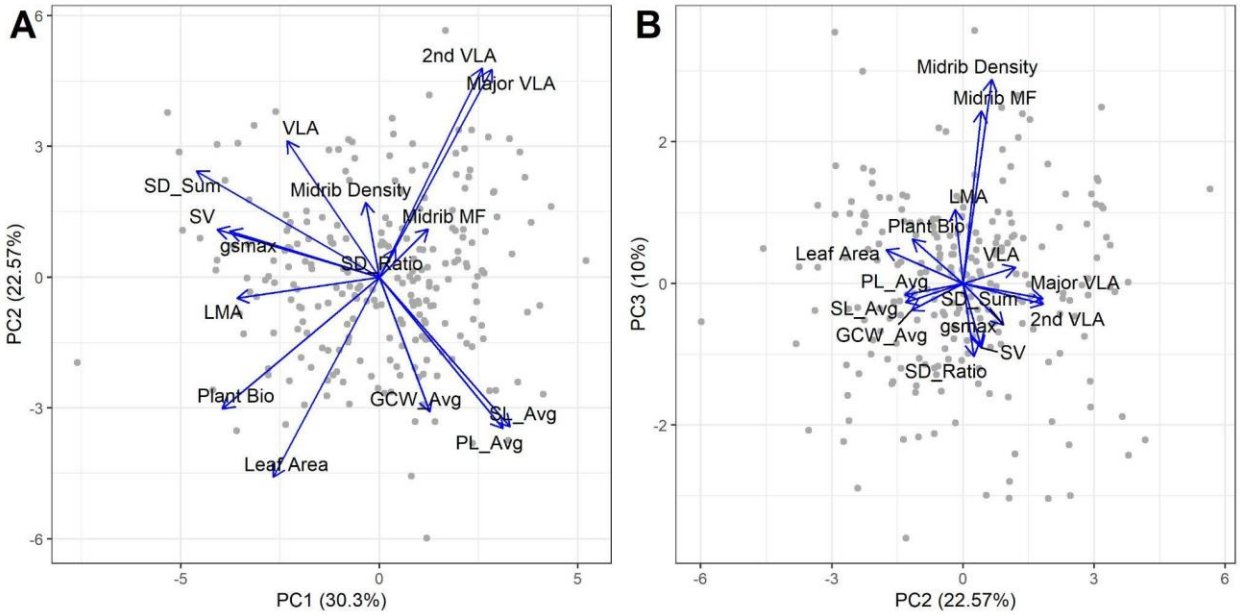

***Figure 4**: Principal component analysis (PCA) of leaf traits using estimated marginal means for each trait. (A) PC1 vs. PC2. (B) PC2 vs. PC3. Traits names that include _Avg (stomatal length [SL], pore length [PL], guard cell width [GCW]) are averages of the values from the top and bottom of the leaf and SD_Sum is the sum of stomatal density from the top and bottom. Trait abbreviations are as defined in Table 1.*

***Table 2:** Trait loadings (fraction of trait variation explained by principal component) of first three principal components (PC). The top three traits per PC are highlighted in bold and with the ranking of the traits for each PC in parentheses. Trait abbreviations are as defined in Table 1.*

Traits	PC1	PC2	PC3
SD_Sum	**16.08 (1)**	4.46 (9)	1.76 (8)
SD_Ratio	0.13 (14)	0.32 (14)	5.48 (4)
LMA	9.75 (5)	0.17 (15)	**5.66 (3)**
Leaf Area	5.42 (9)	**15.94 (3)**	1.16 (9)
Midrib MF	1.14 (13)	0.93 (11)	**30.73 (2)**
Midrib Density	0.09 (15)	2.22 (10)	**42.88 (1)**
GCW_Avg	1.23 (12)	7.23 (7)	0.74 (10)
SL_Avg	8.21 (6)	8.90 (5)	0.36 (12)
PL_Avg	7.38 (7)	9.13 (4)	0.13 (15)
g_{smax}	10.75 (4)	0.84 (13)	3.89 (6)
VLA	4.11 (11)	7.39 (6)	0.26 (13)
2nd VLA	5.10 (10)	**17.40 (1)**	0.44 (11)
Major VLA	6.09 (8)	**17.24 (2)**	0.24 (14)
SV	**12.65 (2)**	0.91 (12)	4.25 (5)
Plant Bio	**11.87 (3)**	6.92 (8)	2.03 (7)

Genetic architecture of observed trait variation

As a first step toward understanding the genetic basis of variation in leaf anatomical traits, we estimated narrow sense heritability for all traits under consideration (Table 3). Midrib traits and VLA (major and minor) had very low heritabilities (0.05 to 0.11) and the confidence intervals of these estimates all overlapped with zero. Stomatal traits had somewhat higher heritability, ranging from 0.08 to 0.20. Heritability values largely reflected our GWA results (see below) in those traits with the lowest heritability estimates had few significant associations.

Our GWA analyses revealed significant associations for 12 out of 24 traits and 2 of the 3 major PC axes (Table 3, Figure 5). There were suggestive associations for several other traits, as well (Figure 5C; Figure S8). Based on observed patterns of LD, we identified a total of 24 independent genomic regions with a significant effect on one or more traits (Figures 5C and S7). Trait co-localization within these regions varied. For example, there were multiple regions on chromosome 11 that associated with both aboveground plant biomass and whole plant biomass, and a region on chromosome 3 (i.e., region 03-01) was significantly associated with g_{smax} while being suggestive for stomatal density and VLA. Interestingly, PC1 (traits primarily related to gas exchange) had a single significant association (on chromosome 12; region 12-01) but no suggestive associations; moreover, this single region was not suggestive for any other traits (Figure 5A). Stomata per vein length (SV), a top contributor to PC1, had a single significant association along with four suggestive regions, none of which corresponded to PC1 (Figure 5B). Similarly, PC3 had a single significant association (on chromosome 3; region 03-02) that did not correspond to any other traits.

Overall, we found somewhat limited evidence for trait colocalization despite the prevalence of significant trait-trait correlations (Figure 5C). Indeed, no region was significantly associated with more than 1-2 traits. However, the inclusion of suggestive associations revealed more support for a common genetic basis. Traits with the largest number of colocalizations were stomatal size and density-related traits. For example, region 13-01 had seven associations including a significant association for stomatal density (top) and suggestive associations for leaf area, pore length (top and average), stomatal density (bottom and sum), and SV (Figure 5C). No significant associations were identified for VLA, although there was a suggestive association in region 03-01, which was significant for g_{smax} and suggestive for stomatal density (both bottom and sum). Allelic effects in this case were consistent with observed trait correlations, with an increase in g_{smax} being associated with an increase in VLA and stomatal density.

Relative effect sizes (RES) for individual associations ranged from 9% to 35%. Summing across regions, traits with significant associations had total RES values of ca. 12-90%. Stomatal density top had the largest amount of variation explained at 89.6% and midrib density the least with 12.3%. Significantly associated regions varied in size, ranging from a single SNP to 6.52 Mbp, though they tended to cluster at the lower end of the range with the majority being < 200 kbp (mean = 592.7 kbp, median = 77.54 kbp; Table S2). These regions contained anywhere from 1 to 135 genes; here again, the significant regions tended to cluster at the lower end of this range (mean = 12.3 genes, median = 2 genes; Table S2). The most gene-rich regions (i.e., 10-

01, 12-01) were significantly associated with variation in plant biomass and PC1, respectively (Figures 5C, S4, and S9). A full list of gene names and annotations is available in Table S2.

Table 3: *List of all traits measured along with estimated narrow sense heritabilities, numbers of significantly associated regions, and total relative effect sizes (RES Sum). Trait abbreviations are as defined in Table 1.*

Trait	h^2	No. Reg.	RES Sum
SD_Bottom (stomata/mm^2)	0.18 (0.10-0.26)	2	45.2
SD_Top (stomata/mm^2)	0.13 (0.07-0.20)	4	89.6
SD_Sum (stomata/mm^2)	0.16 (0.09-0.24)	2	61.3
Stomatal Ratio (bottom/sum)	0.12 (0.05-0.19)	3	53.8
LMA (g/m^2)	0.32 (0.22-0.41)	-	-
Leaf Area (m^2)	0.08 (0.04-0.12)	1	15.4
Midrib MF (g_{midrib}/g_{leaf})	0.05 (-0.01-0.12)	2	20.8
Midrib Density (mg/cm^3)	0.05 (-0.01-0.10)	1	12.3
GCW_Bottom (µm)	0.08 (0.03-0.14)	-	-
GCW_Top (µm)	0.11 (0.05-0.17)	-	-
GCW_Avg (µm)	0.10 (0.04-0.16)	-	-
SL_Bottom (µm)	0.17 (0.09-0.25)	-	-
SL_Top (µm)	0.14 (0.08-0.21)	-	-
SL_Avg (µm)	0.18 (0.10-0.25)	-	-
PL_Bottom (µm)	0.20 (0.12-0.28)	1	22.2
PL_Top (µm)	0.15 (0.08-0.22)	-	-
PL_Avg (µm)	0.19 (0.11-0.27)	-	-
g_{smax} (mol/m^{2}s)	0.27 (0.18-0.36)	2	39.8
VLA (mm/mm^2)	0.06 (0-0.11)	-	-
2nd VLA (mm/mm^2)	0.05 (0.02-0.09)	-	-
Major VLA (mm/mm^2)	0.11 (0.05-0.17)	-	-
SV (stomata/mm)	0.23 (0.15-0.31)	1	19.1
AG Bio (g)	0.49 (0.40-0.58)	3	41.0
Plant Bio (g)	0.46 (0.37-0.56)	6	84.3

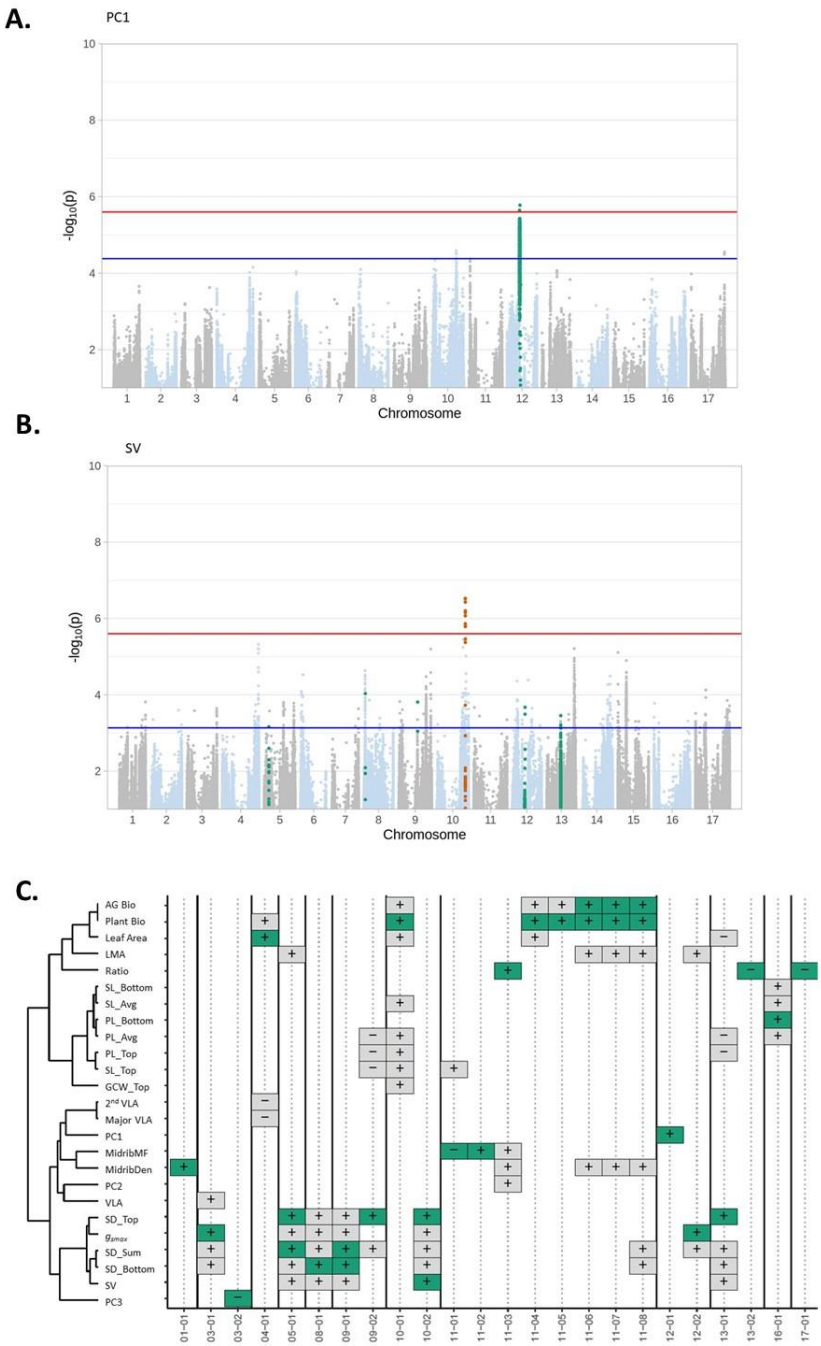

Figure 5: *Examples and summary of GWA results. (A) Manhattan plot of PC1 showing the single significantly associated region on chromosome 12. (B) Manhattan plot of stomata per vein length (SV) showing the single significantly associated region on chromosome 10 and suggestive associations on chromosomes 5, 6, 8, and 13. In both plots, the red line is the significance threshold based on the modified Bonferroni correction and the blue line is the suggestive threshold based on the top 0.1% of all SNPs. Differently colored dots represent all SNPs in a region that are significant or suggestive for at least one trait. (C) Visual summary of all GWA results highlighting colocalization within regions. The dendrogram to the left is based on hierarchical clustering of trait correlations. Green and gray boxes indicate significant or suggestive associations with a given trait, respectively. The sign (+/-) refers to the*

sign of β (the effect of the minor allele on the trait value). Regions are numbered numerically within chromosomes and box/region sizes are arbitrary. Ratio refers to the stomatal ratio. Trait abbreviations are otherwise as defined in Table 1.

Discussion

Stomata and leaf veins are central to plant-water relations and thus potentially important players in determining the performance of plants under water stress. Here, we investigated patterns of variation and covariation in stomatal and vein traits across a diversity panel of inbred cultivated sunflower breeding lines with the goal of improving our understanding of covariation between these traits and their underlying genetic basis. We additionally sought to determine the extent to which observed trait correlations result from genomic co-localization which would indicate that trait relationships are genetically constrained and difficult (if not impossible) to disrupt in the interest of producing novel combinations. In quantifying variation in stomatal and vein traits, we observed numerous correlations amongst traits (Figure 2 and 3). Across traits, we identified three primary axes of variation that we interpreted as being most heavily influenced by traits involved in gas exchange, hydraulics, and leaf construction (Figure 2; Table 2). Subsequently, we performed GWA analyses to examine the genetic architecture of these traits and axes. We found somewhat limited overlap in significant genomic associations across traits, though we did identify significant associations for two of the three multi-trait axes (Figure 5; Table 3).

Patterns of phenotypic variation and trait correlations

Leaf anatomical traits are known to vary widely across species and environments. For example, stomatal length has been shown to vary globally between 10-80 µm with density varying between 5-1000 stomata/mm^2 (Shahinnia *et al.*, 2016; Hetherington and Woodward, 2003). Additionally, a global dataset of 796 species revealed broad variation in estimates of VLA, ranging from 0.1-24.4 mm/mm^2 (Sack and Scoffoni, 2013). In cultivated sunflower, we documented substantial variation in stomatal size and density as well as VLA. Indeed, observed trait values covered 14% (227.5-369.3 stomata/mm^2) of the global range in stomatal density, 18% (26.8-39.4 µm) of the global range in stomatal length, and 23% (7.0-12.6 mm/mm^2) of the global range in VLA (Table 1). While there is a general lack of large datasets describing intraspecific variation in these sorts of traits, particularly for VLA, the ranges that we observed in cultivated sunflower appear quite wide. For example, a global collection of 62 wild accessions of *Arabidopsis thaliana,* grown under benign conditions, covered just 4.3% (17-59 stomata/mm^2) of the global range of stomatal densities (Delgado et al. 2011) while a collection of 330 accessions from across the European range of the same species covered less than 12% (87-204 stomata/mm^2) of the global range (Dittberner et al. 2018). The wide range of variability observed herein is perhaps even more noteworthy given that cultivated sunflower has experienced genetic bottlenecks associated with domestication and improvement that reduced levels of genetic variability as compared to its wild progenitor (Liu and Burke, 2006; Mandel *et al.*, 2011; Park and Burke, 2020).

When analyzed together, leaf anatomical traits exhibited many significant bivariate trait correlations that generally followed expectations based on the literature and their known roles in

plant-water relations (Sack and Scoffoni, 2013; Doheny-Adams *et al.*, 2012; Shahinnia *et al.*, 2016; Figure 2). For example, our data showed that stomatal density and VLA are positively correlated. This was expected as there tends to be a balance between stomata and veins such that water use and carbon acquisition are optimized (Carins Murphy *et al.*, 2014; Brodribb *et al.*, 2007; Sack and Scoffoni, 2013). Additionally, stomatal size and density were negatively correlated, as expected, since the total area allocated to stomata affects total stomatal conductance and thus total photosynthesis (Harrison *et al.*, 2019; Shahinnia *et al.*, 2016; Figure 2) both within (Doheny-Adams *et al.*, 2012) and across species (Hetherington and Woodward, 2003). Overall plant size, estimated as biomass, correlated positively with stomatal density and related traits (e.g., g_{smax} and SV) but negatively with 2nd/major VLA, with larger plants tending to have higher stomatal density and lower 2nd/major VLA. Conversely, stomatal size was unrelated to plant biomass despite its correlation with other leaf traits of interest. Besides scaling with plant mass (Wang *et al.*, 2020), the potential scaling of traits with leaf area is of interest as correlations can arise as a byproduct of trait values scaling with size. Given that stomatal and vein traits were not significantly correlated with leaf area, however, it appears that observed correlations between these traits exist independently of variation in leaf size (Figure 2).

When compared to minor VLA, lower order vein traits (i.e., 2nd and major VLA) exhibited a distinct pattern of trait correlations. These traits were not significantly correlated with any stomatal traits; rather, they exhibited significant correlations with traits related to the investment in leaf production (i.e., LMA and leaf area; Figures 2 and 3). Contrary to expectations based on cross-species comparisons (e.g., Walls, 2011; Kawai and Okada, 2016), 2nd and major VLA were negatively correlated with LMA (Figure 2) suggesting that variation in LMA at this scale may be driven by other, perhaps unmeasured traits such as leaf thickness. Similarly, 2nd and major VLA were negatively correlated with leaf size. This pattern was, however, expected given that major veins are typically formed early in leaf development before being pushed apart as leaf expansion accelerates (Sack and Scoffoni, 2013). In contrast, minor veins are expected to show no such relationship (consistent with our results) because they can be initiated throughout leaf development.

As compared to bivariate analyses, multivariate analyses provide a more holistic view of trait relationships along with possible impacts of leaf anatomical variation on ‘higher level’ traits such as biomass, leaf size, and LMA. When analyzed via PCA, nearly two-thirds of the observed trait variation was captured by the first three PC axes. As noted above, these axes are most heavily influenced by suites of traits involved in gas exchange, hydraulics, and leaf construction. More specifically, plants with lower stomatal density (estimated as stomatal sum) and fewer stomata per vein length (SV), which are primary players in stomatal conductance, tended to be smaller overall. In terms of hydraulic traits, and consistent with the results of our bivariate analyses, plants with greater second and major VLA tended to have smaller leaves. Interestingly, this trait combination is thought to confer greater leaf drought tolerance (Scoffoni *et al.*, 2011). Finally, in terms of leaf construction traits, plants that produced more costly leaves (i.e., leaves with higher LMA) tended to have a greater relative investment in major structural features including midrib density and midrib mass fraction (Figure 4; Table 2), likely reflecting an increase in the mechanical strength of such leaves (Méndez-Alonzo *et al.*, 2013).

Genetic architecture of observed trait variation

Estimates of narrow sense heritability ranged from 0.05 to 0.49 across traits. Notably, vein (including midrib) traits had low heritability estimates while stomatal traits had higher estimates (Table 3). The highest estimates were for biomass-related traits, LMA, and g_{smax} indicating a more substantial contribution of additive genetic effects to observed variation in these traits as compared to others (Kruijer *et al.*, 2015). Somewhat surprisingly given the highly significant effect of genotype on VLA, the heritability estimate for that trait was not significantly different from zero (Tables 1 and 3), indicating little to no contribution of additive genetic effects to observed variation. Midrib density and midrib mass fraction had similarly low heritability estimates, though the evidence of a genotypic effect on the former was less clear, and genotype had no apparent effect on the latter. Interestingly, heritability estimates for the composite trait SV (i.e., stomata per vein length) were noticeably higher than estimates for vein traits alone indicating an additive genetic component of the observed variation in this trait. Collectively, these results suggest that traits with the lowest heritability estimates have limited potential for improvement via breeding while others are likely to be more amenable to such efforts.

Consistent with our heritability estimates, the two biomass-related traits exhibited the largest number of significant associations in our GWA analyses (Table 3; Figure 5C; Figure S8). These traits colocalized with leaf area, midrib density, and LMA but not with any other vein or stomatal traits. Similarly, we identified significant associations for multiple stomatal traits, including stomatal density (bottom, top, and sum) and pore length (bottom), many of which colocalized with suggestive associations (i.e., SNPs with $-\log_{10}(p)$ values in the top 0.1%) for other stomatal traits. Consistent with the low heritability estimates for traits related to vein density (i.e., VLA, 2nd VLA, and major VLA), no significant genomic associations were found for any of these traits. There was, however, one suggestive association for VLA that colocalized with a significant association for g_{smax} (region 03-01) and suggestive associations for stomatal density (bottom and sum), consistent with a presumed functional relationship between these traits. Despite the lack of significant associations for vein-related traits considered on their own, we identified one significant and five suggestive associations for the composite trait SV (i.e., stomata per vein length). These regions tended to colocalize with stomatal traits suggesting that variation in stomatal characteristics is the primary driver of this trait relationship (Figure 5C, Table S1).

Taken together, our trait-by-trait analyses revealed limited evidence for colocalization between stomatal and vein traits despite the existence of widespread and significant correlations between such traits. While this result could be due, at least in part, to the high stringency of our significance threshold and thus the failure to detect true positives – a common challenge in GWA analyses (Gupta *et al.*, 2019) – the identification and inclusion of suggestive regions in our analyses should have helped to mitigate this issue. Nonetheless, observed trait correlations did not appear to be accompanied by clear patterns of genomic colocalization on a single trait basis suggesting a largely independent genetic basis of our traits of interest. However, when multivariate trait relationships were taken into account, our GWA analyses revealed significant associations for two of the three major PC axes (i.e., PC1 and PC3) and a suggestive association for the third (i.e., PC2). For PC1, which is most heavily influenced by traits related to

gas exchange (i.e., stomatal density and SV) along with plant biomass, the single significant association (i.e., region 12-01) colocalized with suggestive associations in 2nd VLA and SV. In contrast, the analysis of PC3, which had midrib density and mass fraction as well as LMA as its top three contributors, identified a novel association (i.e., region 03-02). This region was not identified as being significantly or suggestively associated with any of the individual traits analyzed herein, illustrating the potential value of employing a multivariate approach to GWA analyses (see also, e.g., Yano *et al.*, 2019; Ma *et al.*, 2021). These results also highlight the challenges associated with genetically decoupling certain traits to produce novel phenotypic combinations even though individual trait analyses revealed largely independent genetic architectures.

In terms of effect sizes, the significantly associated regions that we identified tended to have had small to moderate effects (estimated as RES) on trait values. In fact, only three trait/region combinations individually accounted for > 25% of the observed range of trait values across the population (stomatal density [top and sum] in region 05_01 and stomatal density [sum] in region 09_01; Table S1). This result is perhaps not surprising given the relatively low heritability estimates observed for many traits. It is worth noting, however, that the traits with the highest heritability estimates (i.e., biomass-related traits and LMA) tended to be associated with regions of relatively minor effect (i.e., RES < 15%) suggesting that they have a complex genetic basis. In contrast, stomatal traits were associated with some of the largest RES values, suggesting the presence of genes of larger effect and a simpler genetic basis overall. This result was mirrored in the multivariate analyses with the single association underlying PC1 (region 12-01), which is heavily influenced by stomatal traits, accounting for nearly 21% of the observed range of variation in this 'trait' across the population. Unfortunately, most of the genes contained within the significantly associated regions identified herein did not yield obvious candidates for the traits of interest.

While many of the genes in regions of interest were annotated as hypothetical proteins or otherwise showed no clear connection with leaf anatomy, two regions did contain potential genes of interest. Region 10-01, which is significant for total biomass and suggestive aboveground biomass, leaf size, and several stomatal size traits (Figure 5B), includes a gene annotated as a *Putative transcription factor SSXT* (Ha412HOChr10g0435021; GO:0048366 [leaf development]; Table S2). Members of this gene family are thought to play a role in cell size determination in leaves in *Arabidopsis* (Nozaki *et al.*, 2020). This is, however, one of the one of the larger regions that we identified and contains 135 genes so, while this gene appears to be a promising candidate for one or more of the size-related traits that map to this region, this result should be interpreted with caution until functional evidence is available to support its effect on one or more of the associated traits. The other region containing a potential gene of interest, 11-01, contains a *Putative Epidermal Patterning Factor-like protein* (*EPF*; Ha412HOChr11g0479421; GO:0010052 [guard cell differentiation]; Table S2). This region is significant for MidribMF and suggestive for top stomatal length. EPFs are known to be involved in the density of guard cells and epidermal cells (Hara *et al.*, 2009), but how this might relate to midrib mass fraction is unclear. Establishment of a (potential) role for these genes in producing variation in any of the leaf anatomical traits analyzed herein awaits further investigation. Nonetheless, the genomic regions identified during the course of this work, particularly those

with larger effects, represent potential targets for future efforts aimed at modifying leaf anatomical traits in sunflower.

Data Statement

All raw data from the resequencing of the SAM population are stored in the Sequence Read Archive under Bioproject PRJNA353001. SNP set used and genome assembly as in Todesco et al. (2020). Raw phenotypic data and image files will be made available via Dryad upon publication.

Acknowledgements

We thank Kelly Bettinger, Mike Boyd, Kevin Tarner, the rest of the greenhouse staff, and numerous members of the Burke and Donovan labs for help with various aspects of this research. Niki Padgett, Summerlin Courchaine, Kelly Bettinger, Nicole Reisinger, Nadia Kriger, and Jared Bennett provided invaluable assistance with the leaf image processing; Alex Bucksch provided valuable feedback on the automated image analyses; and members of the Burke and Donovan labs provided comments that greatly improved an earlier version of the manuscript. This work was supported by a grant from the NSF Plant Genome Program (IOS-1444522) to JMB as well as funding from the International Consortium on Sunflower Genomics.

References

Bergmann, D.C. and Sack, F.D. (2007) Stomatal development. *Annu. Rev. Plant Biol.*, **58**, 163–181.

Berlyn, G.P., Miksche, J.P., Sass, J.E. and Others (1976) *Botanical microtechnique and cytochemistry*, Iowa State University Press.

Bertolino, L.T., Caine, R.S. and Gray, J.E. (2019) Impact of stomatal density and morphology on water-use efficiency in a changing world. *Front. Plant Sci.*, **10**, 225.

Blackman, B.K., Scascitelli, M., Kane, N.C., Luton, H.H., Rasmussen, D.A., Bye, R.A., Lentz, D.L. and Rieseberg, L.H. (2011) Sunflower domestication alleles support single domestication center in eastern North America. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 14360–14365.

Boer, H.J. de, Eppinga, M.B., Wassen, M.J. and Dekker, S.C. (2012) A critical transition in leaf evolution facilitated the Cretaceous angiosperm revolution. *Nat. Commun.*, **3**, 1221.

Brodribb, T.J., Feild, T.S. and Jordan, G.J. (2007) Leaf maximum photosynthetic rate and venation are linked by hydraulics. *Plant Physiol.*, **144**, 1890–1898.

Buckley, T.N. (2019) How do stomata respond to water status? *New Phytol.* Available at: http://dx.doi.org/10.1111/nph.15899.

Bucksch, A. (2014) A practical introduction to skeletons for the plant sciences. *Applications in Plant Sciences*, **2**. Available at: http://dx.doi.org/10.3732/apps.1400005.

Carins Murphy, M.R., Jordan, G.J. and Brodribb, T.J. (2014) Acclimation to humidity modifies

the link between leaf size and the density of veins and stomata. *Plant Cell Environ.*, **37**, 124–131.

Clevert, D.-A., Unterthiner, T. and Hochreiter, S. (2015) Fast and accurate deep network learning by exponential linear units (ELUs). *arXiv [cs.LG]*. Available at: http://arxiv.org/abs/1511.07289.

Connor, D.J. and Sadras, V.O. (1992) Physiology of yield expression in sunflower. *Field Crops Research*, **30**, 333–389.

Delgado, D., Alonso-Blanco, C., Fenoll, C. and Mena, M. (2011) Natural variation in stomatal abundance of Arabidopsis thaliana includes cryptic diversity for different developmental processes. *Annals of Botany*, **107**, 1247–1258.

Díaz, S., Kattge, J., Cornelissen, J.H.C., et al. (2016) The global spectrum of plant form and function. *Nature*, **529**, 167–171.

Doheny-Adams, T., Hunt, L., Franks, P.J., Beerling, D.J. and Gray, J.E. (2012) Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 547–555.

Dow, G.J., Bergmann, D.C. and Berry, J.A. (2014) An integrated model of stomatal development and leaf physiology. *New Phytologist*, **201**, 1218–1226.

Ellis, B., Daly, D.C., Hickey, L.J., Mitchell, J.D., Johnson, K.R., Wilf, P. and Wing, S.L. (2009) *Manual of Leaf Architecture* 1 edition., Comstock Pub. Associates.

FAO (2018) Food Outlook - Oilcrops. *United Nations*. Available at: http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Oilcrops/Documents/Food_outlook_oilseeds/FO_Oilcrops.pdf [Accessed March 25, 2019].

Faralli, M., Matthews, J. and Lawson, T. (2019) Exploiting natural variation and genetic manipulation of stomatal conductance for crop improvement. *Curr. Opin. Plant Biol.*, **49**, 1–7.

Fiorin, L., Brodribb, T.J. and Anfodillo, T. (2016) Transport efficiency through uniformity: organization of veins and stomata in angiosperm leaves. *New Phytol.*, **209**, 216–227.

Gudesblat, G.E., Schneider-Pizoń, J., Betti, C., et al. (2012) SPEECHLESS integrates brassinosteroid and stomata signalling pathways. *Nat. Cell Biol.*, **14**, 548–554.

Gupta, P.K., Kulwal, P.L. and Jaiswal, V. (2019) Chapter Two - Association mapping in plants in the post-GWAS genomics era. In D. Kumar, ed. *Advances in Genetics*. Academic Press, pp. 75–154.

Hara, K., Yokoo, T., Kajita, R., Onishi, T., Yahata, S., Peterson, K.M., Torii, K.U. and Kakimoto, T. (2009) Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis leaves. *Plant Cell Physiol.*, **50**, 1019–1031.

Harrison, E.L., Arce Cubas, L., Gray, J.E. and Hepworth, C. (2019) The influence of stomatal morphology and distribution on photosynthetic gas exchange. *Plant J.* Available at:

http://dx.doi.org/10.1111/tpj.14560.

Haworth, M., Marino, G., Loreto, F. and Centritto, M. (2021) Integrating stomatal physiology and morphology: evolution of stomatal control and development of future crops. *Oecologia*, **197**, 867–883.

Hetherington, A.M. and Woodward, F.I. (2003) The role of stomata in sensing and driving environmental change. *Nature*, **424**, 901–908.

Hilu, K. and Randall, J. (1984) Convenient method for studying grass leaf epidermis. *Taxon*, **33**, 413–415.

Horikoshi, M. and Tang, Y. (2018) ggfortify: Data Visualization Tools for Statistical Analysis Results. Available at: https://CRAN.R-project.org/package=ggfortify.

Hübner, S., Bercovich, N., Todesco, M., et al. (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat Plants*, **5**, 54–62.

Hussain, M., Farooq, S., Hasan, W., Ul-Allah, S., Tanveer, M., Farooq, M. and Nawaz, A. (2018) Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. *Agric. Water Manage.*, **201**, 152–166.

IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Intergovernmental Panel on Climate Change*.

John, G.P., Scoffoni, C., Buckley, T.N., Villar, R., Poorter, H. and Sack, L. (2017) The anatomical and compositional basis of leaf mass per area. *Ecol. Lett.*, **20**, 412–425.

Kawai, K. and Okada, N. (2016) How are leaf mechanical properties and water-use traits coordinated by vein traits? A case study in Fagaceae J. Watling, ed. *Funct. Ecol.*, **30**, 527–536.

Khan, M.U., Chowdhry, M.A., Khaliq, I. and Ahmad, R. (2003) Morphological response of various genotypes to drought conditions. *Asian J. Plant Sci.*, **2**, 392–394.

Kingma, D.P. and Ba, J. (2014) Adam: A method for stochastic optimization. *arXiv [cs.LG]*. Available at: http://arxiv.org/abs/1412.6980.

Kruijer, W., Boer, M.P., Malosetti, M., Flood, P.J., Engel, B., Kooke, R., Keurentjes, J.J.B. and Eeuwijk, F.A. van (2015) Marker-based estimation of heritability in immortal populations. *Genetics*, **199**, 379–398.

Kruijer, W., Padraic Flood, W. a. C.F.I.W.C.D.C. by and Kooke., R. (2019) heritability: Marker-Based Estimation of Heritability Using Individual Plant or Plot Data. Available at: https://CRAN.R-project.org/package=heritability.

Lei, Z.Y., Han, J.M., Yi, X.P., Zhang, W.F. and Zhang, Y.L. (2018) Coordinated variation between veins and stomata in cotton and its relationship with water-use efficiency under drought stress. *Photosynthetica*, **56**, 1326–1335.

Lenth, R. (2020) emmeans: Estimated Marginal Means, aka Least-Squares Means. Available at: https://CRAN.R-project.org/package=emmeans.

Liu, A. and Burke, J.M. (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*, **173**, 321–330.

Ma, L., Qing, C., Zhang, M., Zou, C., Pan, G. and Shen, Y. (2021) GWAS with a PCA uncovers candidate genes for accumulations of microelements in maize seedlings. *Physiol. Plant.*, **172**, 2170–2180.

Mandel, J.R., Dechaine, J.M., Marek, L.F. and Burke, J.M. (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, Helianthus annuus L. *Theor. Appl. Genet.*, **123**, 693–704.

Mandel, J.R., Nambeesan, S., Bowers, J.E., Marek, L.F., Ebert, D., Rieseberg, L.H., Knapp, S.J. and Burke, J.M. (2013) Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.*, **9**, e1003378.

Masalia, R.R., Temme, A.A., Torralba, N. de L. and Burke, J.M. (2018) Multiple genomic regions influence root morphology and seedling growth in cultivated sunflower (Helianthus annuus L.) under well-watered and water-limited conditions. *PLoS One*, **13**, e0204279.

Méndez-Alonzo, R., Ewers, F.W. and Sack, L. (2013) Ecological variation in leaf biomechanics and its scaling with tissue structure across three mediterranean-climate plant communities. *Funct. Ecol.*, **27**, 544–554.

Muir, C.D. (2018) Light and growth form interact to shape stomatal ratio among British angiosperms. *New Phytol.*, **218**, 242–252.

Muir, C.D., Pease, J.B. and Moyle, L.C. (2014) Quantitative genetic analysis indicates natural selection on leaf phenotypes across wild tomato species (Solanum sect. Lycopersicon; Solanaceae). *Genetics*, **198**, 1629–1643.

NOAA Drought: Monitoring Economic, Environmental, and Social Impacts. *NOAA: National Centers for Environmental Information*. Available at: https://www.ncdc.noaa.gov/news/drought-monitoring-economic-environmental-and-social-impacts [Accessed May 11, 2021].

Nozaki, M., Kawade, K., Horiguchi, G. and Tsukaya, H. (2020) an3-Mediated Compensation Is Dependent on a Cell-Autonomous Mechanism in Leaf Epidermal Tissue. *Plant Cell Physiol.*, **61**, 1181–1190.

Park, B. and Burke, J.M. (2020) Phylogeography and the Evolutionary History of Sunflower (Helianthus annuus L.): Wild Diversity and the Dynamics of Domestication. *Genes* , **11**. Available at: http://dx.doi.org/10.3390/genes11030266.

Paszke, A., Gross, S., Chintala, S., et al. (2017) Automatic differentiation in PyTorch. Available at: https://openreview.net/pdf?id=BJJsrmfCZ [Accessed September 9, 2020].

Poorter, H., Niinemets, Ü., Poorter, L. and Wright, I.J. (2009) Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New*. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2009.02830.x.

R Core Team (2013) *R: The R project for statistical computing*, Available at: https://www.r-project.org/ [Accessed April 17, 2019].

Reich, P.B. (2014) The world-wide “fast–slow”plant economics spectrum: a traits manifesto. *The Journal of Ecology*. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2745.12211.

Richardson, F., Jordan, G.J. and Brodribb, T.J. (2020) Leaf hydraulic conductance is linked to leaf symmetry in bifacial, amphistomatic leaves of Sunflower. *J. Exp. Bot.* Available at: http://dx.doi.org/10.1093/jxb/eraa035.

Ries, L.L., Purcell, L.C., Carter, T.E., Edwards, J.T. and King, C.A. (2012) Physiological Traits Contributing to Differential Canopy Wilting in Soybean under Drought. *Crop Sci.*, **52**, 272–281.

Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, pp. 234–241.

Sack, L. and Scoffoni, C. (2013) Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. *New Phytologist*, **198**, 983–1000.

Schindelin, J., Arganda-Carreras, I., Frise, E., et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, **9**, 676–682.

Scoffoni, C., Rawls, M., McKown, A., Cochard, H. and Sack, L. (2011) Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. *Plant Physiol.*, **156**, 832–843.

Scoffoni, C. and Sack, L. (2013) Quantifying leaf vein traits. *Prometheus Wiki*. Available at: http://prometheuswiki.org/tiki-index.php?page=Quantifying+leaf+vein+traits&highlight=leaf%20clearing [Accessed March 21, 2019].

Shahinnia, F., Le Roy, J., Laborde, B., Sznajder, B., Kalambettu, P., Mahjourimajd, S., Tilbrook, J. and Fleury, D. (2016) Genetic association of stomatal traits and yield in wheat grown in low rainfall environments. *BMC Plant Biology*, **16**, 150.

Shi, P., Jiao, Y., Diggle, P.J., Turner, R., Wang, R. and Niinemets, Ü. (2021) Spatial distribution characteristics of stomata at the areole level in Michelia cavaleriei var. platypetala (Magnoliaceae). *Ann. Bot.*, **128**, 875–886.

Tang, Y., Horikoshi, M. and Li, W. (2016) ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *The R Journal*, **8**. Available at: https://journal.r-project.org/.

Temme, A.A., Kerr, K.L., Masalia, R.R., Burke, J.M. and Donovan, L.A. (2020) Key Traits and Genes Associate with Salinity Tolerance Independent from Vigor in Cultivated Sunflower. *Plant Physiol.*, **184**, 865–880.

Todesco, M., Owens, G.L., Bercovich, N., et al. (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, **584**, 602–607.

Vasco, A., Thadeo, M., Conover, M. and Daly, D.C. (2014) Preparation of samples for leaf architecture studies, a method for mounting cleared leaves. *Appl. Plant Sci.*, **2**. Available at: http://dx.doi.org/10.3732/apps.1400038.

Walls, R.L. (2011) Angiosperm leaf vein patterns are linked to leaf functions in a global-scale data set. *American Journal of Botany*, **98**, 244–253.

Wang, Y., Donovan, L.A. and Temme, A.A. (2020) Plasticity and the role of mass-scaling in allocation, morphology, and anatomical trait responses to above- and belowground resource limitation in cultivated sunflower (Helianthus annuus L.). *Plant Direct*, **4**, e02924.

Wei, T. and Simko, V. (2017) *R package "corrplot": Visualization of a Correlation Matrix*, Available at: https://github.com/taiyun/corrplot.

Weyers, J.D.B. and Johansen, L.G. (1985) Accurate Estimation of Stomatal Aperture from Silicone Rubber Impressions. *New Phytol.*, **101**, 109–115.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Available at: https://ggplot2.tidyverse.org.

Wills, D.M. and Burke, J.M. (2006) Chloroplast DNA variation confirms a single origin of domesticated sunflower (Helianthus annuus L.). *J. Hered.*, **97**, 403–408.

Wright, I.J., Reich, P.B., Westoby, M., et al. (2004) The worldwide leaf economics spectrum. *Nature*, **428**, 821–827.

Xing, K., Niinemets, Ü., Rengel, Z., Onoda, Y., Xia, J., Chen, H.Y.H., Zhao, M., Han, W. and Li, H. (2021) Global patterns of leaf construction traits and their covariation along climate and soil environmental gradients. *New Phytol.*, **232**, 1648–1660.

Xiong, D. and Flexas, J. (2020) From one side to two sides: the effects of stomatal distribution on photosynthesis. *New Phytol.*, **228**, 1754–1766.

Xu, H., Blonder, B., Jodra, M., Malhi, Y. and Fricker, M. (2020) Automated and accurate segmentation of leaf venation networks via deep learning. *New Phytol.* Available at: http://dx.doi.org/10.1111/nph.16923.

Yano, K., Morinaka, Y., Wang, F., et al. (2019) GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 21262–21267.

Yu, F. and Koltun, V. (2015) Multi-scale context aggregation by dilated convolutions. *arXiv [cs.CV]*. Available at: http://arxiv.org/abs/1511.07122.

Zhang, S.-B., Guan, Z.-J., Sun, M., Zhang, J.-J., Cao, K.-F. and Hu, H. (2012) Evolutionary association of stomatal traits with leaf vein density in Paphiopedilum, Orchidaceae. *PLoS One*, **7**, e40080.

Zhao, W.-L., Siddiq, Z., Fu, P.-L., Zhang, J.-L. and Cao, K.-F. (2017) Stable stomatal number per minor vein length indicates the coordination between leaf water supply and demand in three leguminous species. *Sci. Rep.*, **7**, 2211.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. and Weir, B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association

studies. *Nat. Genet.*, **44**, 821–824.

Supporting Information:

Figure S1: *Comparison of human measured and neural network estimated vein lengths on test-set images (N=85).*

Figure S2: *Network architecture used in-place of the U-Net structure after the last max pool layer through the first up convolution. The 32x32x512 input tensor represents the output of the last max-pool layer. The 56x56x512 output tensor is the same as the output size of the first up convolution. + represents element-wise addition of the output tensors.*

Figure S3*: Cross-validated training (red) and validation (black) loss of model with final hyperparameters. Shading indicates standard error.*

Figure S4: *Bivariate plots for all trait correlations presented in Figure 2. Points represent the estimated marginal means of genotypes. Blue line represents the fitted regression line. Units for all traits are as in Table 1.*

Figure S5: *Principal component analysis (PCA) of all measured leaf traits using estimated marginal means for each trait. For stomatal density, length, pore length, and guard cell width, trait values are included separately for the top and bottom of the leaf.*

Figure S6: *Scree plot showing principal components vs. percent variance explained.*

Figure S7. *Linkage disequilibrium heatmaps for all significant SNPs found on each chromosome. Linkage disequilibrium was estimated as R^2 between all significant SNPs per chromosome. Colored blocks represent haplotypic blocks that the SNPs belong to based on the full haplotype map (Figure S10). Block membership is shown for both an analysis based on the full, genome-wide collection of SNPs ("genome") as well as a re-analysis based on only significant SNPs ("significant"). Colors are arbitrary. For full methodological details, see Temme et al. (2020).*

Figure S8: *Manhattan plots resulting from GWA analyses for all traits. The red line is the significance threshold based on the modified Bonferroni correction (see text for details) and the blue line is the suggestive threshold based on (i.e., top 0.1% of all SNPs). Colored dots represent SNPs that are significant or suggestive for at least one trait. Color of dots is arbitrary.*

Figure S9: *Distribution of the number of genes per significant region. Points represent the number of genes in each significant block. Note the log scale on the y-axis.*

Figure S10: Visualization of *haplotypic blocks across the sunflower genome. Blocks are shown for each of the 17 sunflower chromosomes. Blocks are colored in an alternating fashion and colors are arbitrary. Black dots on the x-axis indicate singleton SNPs that did not fall within blocks.*

Table S1: *All significant and suggestive regions underlying observed trait variation. Suggestive regions are defined as containing SNPs in the top 0.1% of all SNPs that co-localizing with a significant region.*

Region boundaries were determined by re-running the haplotype blocking process using SNPs identified as being significant based on the GWA analyses. Haplotype block refers to the unique block in the genome based on the full set of SNPs and NSNP gives the number of significant SNPs in that haplotype block. Beta refers to the sign and magnitude of the effect of the minor allele and RES is the relative effect size of the SNPs in the haplotype block.

Table S2: *List of genes per region along with significant/suggestive trait associations. Gene names come from annotation of the sunflower HA412-HOv2 genome assembly. Physical size of haplotype blocks is listed.*

Methods S1: *Neural Network Architecture, Training and Prediction.* The neural network followed that of the U-Net (Ronneberger *et al.*, 2015) deep neural network. High-dilation convolutions (Yu and Koltun, 2015) were integrated by replacing the U-Net architecture after the last max-pool through the first up convolution with the structure shown in Figure S1. In addition, all RELU activations were replaced with ELUs (Clevert *et al.*, 2015). The network architecture takes input images of size 572x572x3 pixels and outputs a segmented target image of 388x388x1 pixels that corresponds to the center of the 572x572 image. For training, input images were created by randomly selecting 388x388x3 regions from the leaf images along with the surrounding 92 pixels to create the 572x572x3 input image. As in Ronneberger et al. (2015), when surrounding data was missing at the edge of the image, the pixels were extrapolated by mirroring. The 572x572x3 bright-field leaf input images were independently normalized to have a mean of 0 and a standard deviation of 1. The network was coded using the pytorch (Paszke *et al.*, 2017) framework and trained using backpropagation using the Adam optimizer (Kingma and Ba, 2014). The loss function was weighted binary cross entropy. Non-vein pixels were given a weight of 0.05 and in-vein pixels a weight of 1. The weighting of the non-vein pixels was chosen using 3-fold cross validation of the training set of images using a grid search of 0.01, 0.05, 0.1, and 0.2 for the weight. Cross-validation training was performed out to 3000 batches, but there was no further decrease in the loss on the validation sets past 1900 batches (Figure S3).

Attempts at stitching 388x388 segmented tiles together to recreate the full size bright-field images resulted in incorrect segmentations at the boundaries between tiles. To avoid the boundary effects, the full-sized images (2584x1936 pixels) were mirror padded to 2876 x 2300 and passed through the network as one input. The images were normalized as described for the image tiles used in training. The resulting segmented images were cropped back to 2584x1936 pixels to remove padding pixels. The padding was used to both avoid edge effects – as done during network training – and to ensure all the convolution operations within the network were valid. The network outputs the probability a pixel is within a vein. The in-vein segmentation cutoff of 0.2 maximized the correlation between the number of pixels in the hand drawn vein lines and in that extracted with the network in a 3-way cross validation grid search of values spanning 0.2 to 0.9 with 0.1 increments.