Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PhenoFrame: Phenotyping Plant Descriptor and Software Suite

Jorge Vazquez
Purdue University
Karla Cuellar-Perez
University of Nebraska-Lincoln
Jensina Davis
University of Nebraska-Lincoln
Zhongjie Ji
University of Nebraska-Lincoln
Sumeet Mankar
University of Nebraska-Lincoln
James Schnable
University of Nebraska-Lincoln
Bedrich Benes
bbenes@purdue.edu

Purdue University

Article

Keywords: Plant phenotyping, Canopy architecture, Procedural modeling, Data standards, Quantitative genetics, Sorghum, Maize

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11381357/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. JS has equity interests in Data2Bio and Dryland Genetics and has performed paid work for Alphabet. The authors declare no other competing interests.

PhenoFrame: Phenotyping Plant Descriptor and Software Suite

Jorge Vazquez[1], Karla M. Cuéllar-Pérez[2], Jensina M. Davis[2], Zhongjie Ji[2], Sumeet P. Mankar[2], James Schnable[2], Bedrich Benes[1*]

[1]Department of Computer Science, Purdue University, 305 N University St, West Lafayette, 47907, Indiana, USA.
[2]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, 1825 N 38th St, Lincoln, 68583, Nebraska, USA.

*Corresponding author(s). E-mail(s): bbenes@purdue.edu;
Contributing authors: vazque81@purdue.edu; kcuellarperez2@huskers.unl.edu; jdavis132@huskers.unl.edu; zji7@unl.edu; smankar2@unl.edu; schnable@unl.edu;

Abstract

Exchanging three-dimensional plant reconstructions between measurement and modeling workflows presents significant challenges. PhenoFrame is introduced as a shared descriptor and software suite that offers both a skeleton mode, utilizing observed leaf centerlines, and a procedural mode, in which each centerline is replaced by fitted parameters. A total of **324** sorghum plants and **3,317** repeated scans of **80** maize plants were converted into both modes. For sorghum, the median descriptor size was reduced from **9.6** to **4.5** kB per plant, with a median mode-to-mode Chamfer distance of **0.65** cm (**0.46** cm for maize). For **9** of **13** sorghum traits, procedural-mode estimates reached a broad-sense heritability of at least **0.20** (**10** of **13** for voxel-derived traits), as did all **13** maize traits. A descriptor-based stochastic generator approximated the marginal trait distributions, overstating the mean absolute between-node leaf-angle correlation (**0.74** versus **0.25** in actual plants). Genome-wide association analysis of the descriptor azimuth traits identified several candidate genes associated with auxin signaling, which is expected to regulate phyllotaxy.

Keywords: Plant phenotyping, Canopy architecture, Procedural modeling, Data standards, Quantitative genetics, Sorghum, Maize

1 Introduction

Feeding a projected global population of 9.7 billion by 2050 will require approximately 50% more food than was produced in 2012 [1]. However, yields of major crops such as maize, rice, soybean, and wheat are increasing by only 0.9 to 1.6% annually, which is below the 2.4% needed to double production [2]. Bridging this gap requires improved yields and a deeper understanding of the genotype-phenotype relationship. The most influential traits for yield, such as plant canopy architecture (including the spatial arrangement of leaves, stem internodes, and reproductive organs), are often the most challenging to measure at scale. In maize and sorghum, architectural features such as leaf angle, phyllotaxy, and internode length directly influence light capture and distribution, thereby affecting yield [3, 4]. This relationship is exemplified by the maize *lac1* mutant, which optimizes leaf angle layer by layer; upper leaves become more upright and gradually flatten toward the base, increasing grain yield in large-scale field trials under high planting densities [5]. Identifying the genetic basis of these traits requires phenotyping thousands of individual plants. While modern sensing technologies have made large-scale phenotyping feasible, computational assistance remains essential. Canopy architecture is inherently three-dimensional (3D), yet most traditional methods reduce it to two-dimensional (2D) proxies, losing critical spatial information that only 3D reconstruction can recover [4]. Manual measurement exacerbates this challenge: quantifying phyllotaxy alone requires 10 to 20 minutes per plant [3], rendering large-scale phenotyping and genome-wide association studies impractical without computational support. As genomic data continues to outpace phenotypic data [6], computational frameworks that generate accurate 3D representations of plant architecture offer a scalable solution for trait extraction, genetic mapping, and simulating emergent canopy behaviors that are otherwise difficult to measure experimentally [7].

Advancements in understanding the genetic determinants of canopy structure in maize and sorghum have been achieved through quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS). In maize, GWAS conducted on the nested association mapping (NAM) population demonstrated that leaf angle and other architectural traits are controlled by many loci of small effect, including variants at the *liguleless* genes [8]. Comparable studies in sorghum, particularly those utilizing the global sorghum NAM, have elucidated the genetic architecture of inflorescence morphology and structural traits with a resolution similar to that achieved in maize [9]. Analytical approaches in this field commonly employ established software suites, including TASSEL [10], GAPIT [11], and multi-locus mixed models such as FarmCPU [12] and BLINK [13]. More recent Python-based implementations, such as PANICLE [14], have integrated these methods within the scientific Python ecosystem, offering a nearly standardized workflow from genotypes and covariates to marker-trait associations.

Yet standardization ends with the analysis interface: model selection, covariate structure, and preprocessing decisions remain study-specific, and the phenotypic inputs used in these pipelines are not uniform or easily reproducible across laboratories. Phenotyping capacity has lagged behind sequencing capacity for over a decade [15, 16]. Although high-throughput imaging platforms have narrowed this gap [17, 18], their outputs are typically formatted in pipeline-specific representations,

complicating the comparison of derived traits across studies. Architectural traits, which are inherently 3D and require dozens of derived quantities per plant, are particularly exposed to this bottleneck: the absence of a shared, biologically interpretable phenotypic representation limits both the reproducibility and the statistical power of genetic mapping.

Various phenotypic representations arise from computational pipelines used to model and reconstruct plant architecture. Current methods include procedural modeling, 3D reconstruction, and skeletonization. Procedural modeling work focuses on generating geometry based on a series of rules. A major method involves encoding these rules as parallel string-rewriting systems, known as Lindenmayer systems (L-systems) [19], which enable the simulation of nutrient and signaling processes in plants [20], as well as the generation of complex 3D geometries and interactions with the environment [21, 22], producing the geometry of individual plants based on a species-specific model. Other studies have developed a general plant model that is then either adapted to other plant species [23] or applied to an entire field in order to achieve biologically plausible results [24].

Alongside procedural modeling (which requires substantial effort to tune for each species), work has aimed to obtain a skeletonized version of the plant from 3D point clouds, enabling accurate extraction of phenotypic traits [25]. Furthermore, current research aims to achieve complete plant reconstruction, either by applying shape fitting to segmented maize point clouds [26], by using voxel carving on sorghum point clouds [4], or by employing new deep learning methods to derive procedural rules for tree generation [27].

Moreover, frameworks for the modeling and simulation of plant architecture have been created, such as L-Py [28], which permits the modeling of plant structures using L-systems, OpenAlea [29], which offers a visual and interactive interface for the development of functional–structural plant model (FSPM) simulations, and AMAP-studio [30], which provides a suite designed for botanists and agronomists and allows plant architectures to be simulated at various scales based on a multiscale tree graph (MTG), a data structure introduced in [31]. Other work has focused either on developing a generalizable plant structure for efficient modeling or on creating highly specific structures for detailed, species-specific modeling [32]. The variety of these approaches justifies the need for a descriptor that explicitly records plant geometry and derived traits, so we can measure the effects of conversion between representations.

These methods are constrained by challenges in developing high-throughput algorithms, reconstructing plants through development, and analyzing plant function, including extracting phenotypic traits for plant science applications [33]. FSPMs often lack collaborative development among plant scientists, mathematicians, modelers, and computer scientists. They also present a steep learning curve for new users [34], and there is currently no generalized, interpretable plant representation that facilitates straightforward trait interpretation, can be generated by reconstruction methods, and supports the creation of large plant datasets.

In other scientific domains, shared data acquisition and analysis tools have established common representations that serve as general formats for generation, modeling,

and analysis. For example, in biological imaging, the Open Microscopy Environment (OME) defines an XML data format that enables visualization and analysis using a unified representation [35]. In computational biology, the Systems Biology Markup Language (SBML) allows simulators and analysis tools to operate on the same model [36]. In neurophysiology, Neurodata Without Borders (NWB) integrates diverse data types into a single format accessible to an ecosystem of analysis tools [37]. In plant science, shareable software suites are an emerging trend [38].

To provide an interpretable representation of grass canopy architecture, this work introduces PhenoFrame, a hierarchical, human-readable descriptor and an accompanying software suite. The schema records each leaf either as a set of observed skeleton control points ($\mathcal{S}$) or as a procedurally fitted representation defined by its trait values ($\mathcal{P}$). Both approaches utilize a unified geometry and trait-processing pipeline. Previously published sorghum skeletons and repeatedly reconstructed maize data are converted into both formats, enabling assessment of storage requirements, geometric fidelity, agreement between reference and derived traits, and heritable signal for matched plants and leaves. The descriptor's utility is further demonstrated through synthetic generation and association analysis, and its limitations are explicitly evaluated.

2 Results

PhenoFrame was evaluated through descriptor-to-measurement consistency, skeleton and procedural reconstructions, heritability, stochastic generation, and an exploratory genome-wide association study (GWAS). The primary comparison utilized the same plants and retained leaves in both descriptor modes. The sorghum dataset comprised 324 reconstructed plants, while the maize geometry and trait analyses pooled 3,317 repeated scans from 80 plants. These experiments assess the benefits and limitations of compact procedural encoding, rather than establishing general accuracy across all grass species.

2.1 Terminology

Representations: PhenoFrame uses one XML descriptor $\mathcal{D}$ with two leaf modes. Skeleton mode $\mathcal{S}$ stores observed centerline control points; procedural mode $\mathcal{P}$ stores trait values (fitted angle, droopiness, and length) and regenerates the centerline. Both are populated from an extracted voxel skeleton. The original voxel grid is a geometric comparator, while PCA-based leaf traits provide reference values (see Figure 1).

Differences: We denote a difference in trait values or geometry by Δ, with a subscript naming the compared sources; for example, $\Delta_{\mathcal{S}\leftrightarrow\mathcal{P}}$ denotes the Chamfer distance between the two descriptor modes.

2.2 PhenoFrame descriptor is self-consistent

The initial verification assessed whether the procedural model recovers values declared in $\mathcal{D}$ after the trait pipeline generates geometry and measures it. For $N = 500$ generated descriptors comprising $6,213$ leaves, the median absolute deviations were 0.002

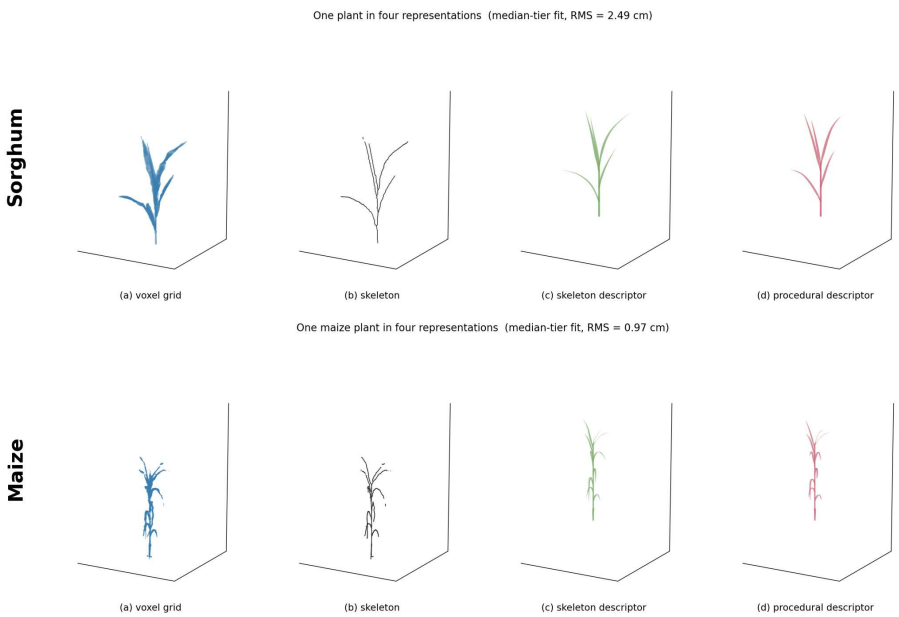


Fig. 1: Sorghum and maize plants shown in various formats: (a) Voxel Reconstruction, (b) Extracted Skeleton, (c) skeleton descriptor $\mathcal{S}$, and (d) the fitted three-parameter procedural descriptor $\mathcal{P}$.

mm for leaf length, 0.10 mm for cumulative internode distance, 0.069° for leaf azimuth (excluding near-vertical leaves), and 1.05° for leaf angle (Figure 2). These medians indicate close agreement for most generated leaves.

2.3 PhenoFrame descriptor captures real plant architecture

We used extracted leaf skeletons to populate $\mathcal{S}$ and fit $\mathcal{P}$ for 324 sorghum plants from [4] and 3,317 repeated scans of 80 maize plants (see Section 4.2.1). We used the published leaf angle and azimuth values of Davis et al. [3] as the sorghum reference, and we reimplemented the same method for maize. Following the published protocol, we compared the same retained lower-four leaves in both descriptor modes. These comparisons are conditional on the leaf-quality and procedural-fit filters described in Section 4.2.1.

On the matched sorghum lower-four leaves, skeleton-mode agreement with the PCA-based reference was $r^2(\theta) = 0.46$ and $r^2(\phi) = 0.53$; the corresponding pooled maize values were 0.49 and 0.43 (Table 1 and Figure S1). These are comparisons between algorithms acting on the same reconstructions, not independent physical ground truth.

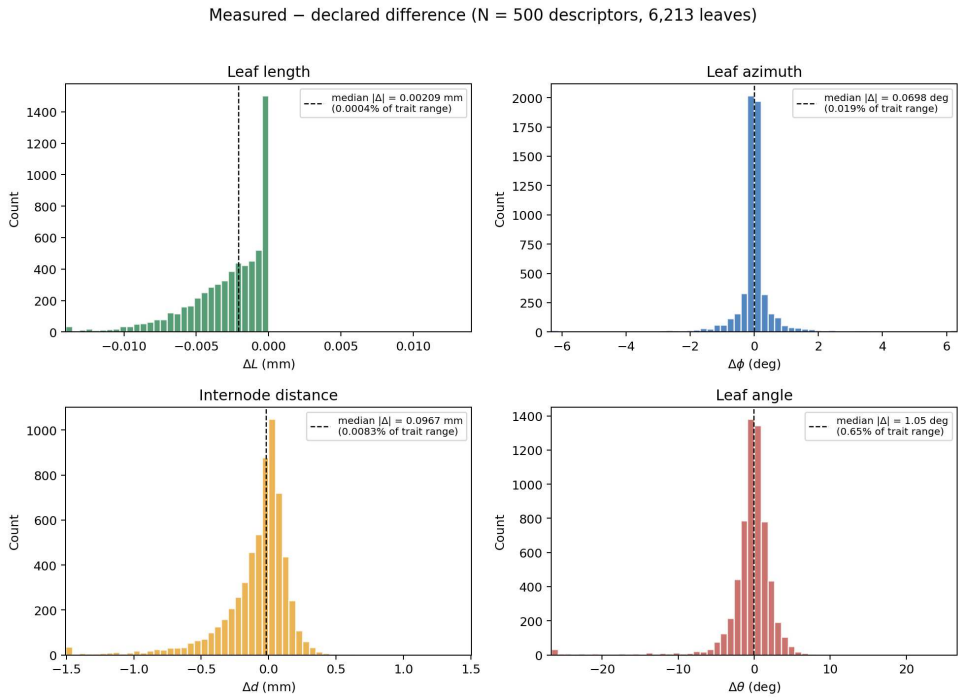


Fig. 2: We generate the geometry from the descriptor $\mathcal{D}$, measure the generated internode length, leaf length, angle, and azimuth, and compare them to the input values in $\mathcal{D}$.

Species	Trait	r	r^2	Median absolute error
Sorghum	Leaf angle (θ)	0.68	0.46	4.90°
	Leaf azimuth (ϕ)	0.73	0.53	5.29°
Maize	Leaf angle (θ)	0.70	0.49	4.71°
	Leaf azimuth (ϕ)	0.65	0.43	4.41°

Table 1: Skeleton-mode $\mathcal{S}$ agreement with reference leaf angle and azimuth traits on the same retained lower-four leaves. Sorghum reference values were published by Davis et al. [3]; we reimplemented the method for maize. Maize values pool repeated scans from 80 plants.

2.4 Procedural model reproduces real geometry

A leaf can be represented by the explicit control points of $\mathcal{S}$ or by the fitted angle, droopiness, and length parameters of $\mathcal{P}$ (Table S1). $\mathcal{P}$ was fitted to the same reconstructed skeletons, and whole-plant geometry was compared on matched instances. The median Chamfer distance between modes was 0.65 cm in sorghum and 0.46 cm in maize. In sorghum, the median voxel-to-mesh Chamfer distance was 3.58 cm for $\mathcal{S}$

and 3.82 cm for $\mathcal{P}$; the corresponding Hausdorff distances were 18.04 and 20.04 cm (Figure 3 and Table 2).

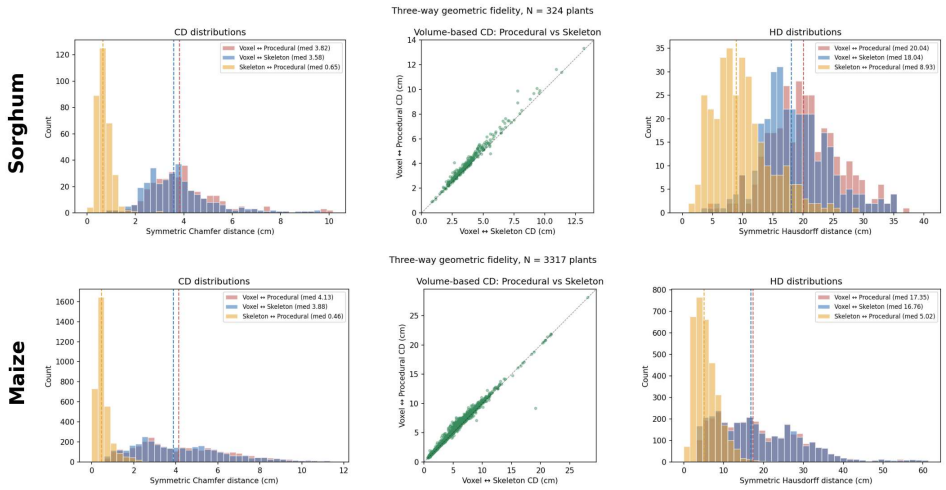


Fig. 3: Chamfer (CD) and Hausdorff (HD) distances for skeleton and procedural descriptor modes and the voxel reconstructions. Voxel-to-mesh distances are conditional on a fixed blade width.

	Sorghum		Maize	
Quantity	Skeleton $\mathcal{S}$	Procedural $\mathcal{P}$	Skeleton $\mathcal{S}$	Procedural $\mathcal{P}$
Leaves, lower four (both modes)	1,168		11,301	
Median descriptor size (kB per plant)	9.6	4.5	3.9	1.9
$r(\theta)$	0.68	0.60	0.70	0.60
$r^2(\theta)$, lower four	0.46	0.36	0.49	0.36
Median $\lvert\Delta\theta\rvert$ vs PCA reference (°)	4.90	5.36	4.71	8.39
$r(\phi)$	0.73	0.68	0.65	0.64
Circular $r_{cc}(\phi)$	0.23	0.32	0.79	0.80
$r^2(\phi)$, lower four	0.53	0.47	0.43	0.41
Median $\lvert\Delta\phi\rvert$ vs PCA reference (°)	5.29	6.89	4.41	5.02
Median mode-to-mode Chamfer distance (cm)	0.65		0.46	
Median Chamfer distance to voxel grid (cm)	3.58	3.82	3.88	4.13
Median Hausdorff distance to voxel grid (cm)	18.04	20.04	16.76	17.35

Table 2: Compression and trait trade-offs between skeleton mode $\mathcal{S}$ and procedural mode $\mathcal{P}$. Sorghum and maize sizes are median serialized descriptor sizes. Trait rows compare the same retained lower-four leaves with reference values derived from the same voxel skeletons. Geometric distances compare each reconstructed instance with the voxel grid, except the mode-to-mode row.

For the same retained sorghum leaves, the procedural representation reduced leaf-angle r^2 from 0.46 to 0.36 and increased the median absolute leaf-angle error from 4.90° to 5.36°; azimuth r^2 decreased from 0.53 to 0.47. In maize, the median leaf-angle error increased from 4.71° to 8.39° (Table 2). The fitted parameters continue to vary coherently with leaf rank (Figure S2), but the three-parameter curve cannot capture every feature of an observed centerline. Procedural mode is suitable for compact population summaries, whereas skeleton mode retains more detailed per-leaf geometry.

Conversion of the maize scans to $\mathcal{S}$ and $\mathcal{P}$ demonstrates a similar geometry-versus-trait trade-off (Table 2 and Figure 3). The median mode-to-mode Chamfer distance was 0.46 cm, but the procedural mode exhibited a larger leaf-angle error.

2.5 PhenoFrame's representations preserve heritable signal

Broad-sense heritability was estimated for 13 lower-canopy traits from the voxel pipeline, $\mathcal{S}$, and $\mathcal{P}$ using the same 128 sorghum plants from 44 replicated genotypes and 80 maize plants from 20 genotypes. The model used was $H^2 = \sigma_G^2/(\sigma_G^2 + \sigma_e^2/n)$, with $n = 2$ sorghum and $n = 4$ maize replicates. Trait-source comparisons utilized common retained leaves and, for divergence angles, only consecutive-leaf pairs retained by all three sources.

The 13 traits comprise the per-node leaf inclination $\theta_0, \ldots, \theta_4$ of the lowest five extant leaves, the per-node deviation from 180° of the divergence angle between consecutive leaves of that set, $\Phi_1, \ldots, \Phi_4$ (as defined for Φ_t in Table S2), and four per-plant aggregates (the mean and median of θ and of Φ over the lower five leaves). All 13 are derived from the per-leaf traits of Table S2 and use the same definitions across the three sources. Divergence-angle estimates also depend on the common-pair filter, which retains 59.1% of sorghum and 59.5% of maize pairs.

Species	Source	Traits $H^2 \geq 0.20$	Median H^2	Mean H^2
Sorghum ($n = 2$)	Voxel	10 / 13	0.35	0.35
Sorghum ($n = 2$)	Skeleton $\mathcal{S}$	9 / 13	0.39	0.35
Sorghum ($n = 2$)	Procedural $\mathcal{P}$	9 / 13	0.48	0.41
Maize ($n = 4$)	Voxel	13 / 13	0.87	0.76
Maize ($n = 4$)	Skeleton $\mathcal{S}$	12 / 13	0.91	0.74
Maize ($n = 4$)	Procedural $\mathcal{P}$	13 / 13	0.90	0.73

Table 3: Broad-sense heritability (REML) of 13 lower-canopy traits on leaves retained by both descriptor modes, with divergence-angle traits calculated on pairs retained by all three sources. n is the number of replicates per genotype. In sorghum, the procedural mode has the highest median point estimate, driven by inclination traits, but not the most traits above threshold. Sorghum: 128 plants, 44 genotypes. Maize: 80 plants, 20 genotypes.

In sorghum, median H^2 across the 13 traits was 0.35 for the voxel source, 0.39 for $\mathcal{S}$ and 0.48 for $\mathcal{P}$; the numbers at or above $H^2 = 0.20$ were 10, 9 and 9, respectively

(Figure 4 and Table 3). The higher procedural median was driven by leaf inclination, while divergence-angle heritability generally weakened.

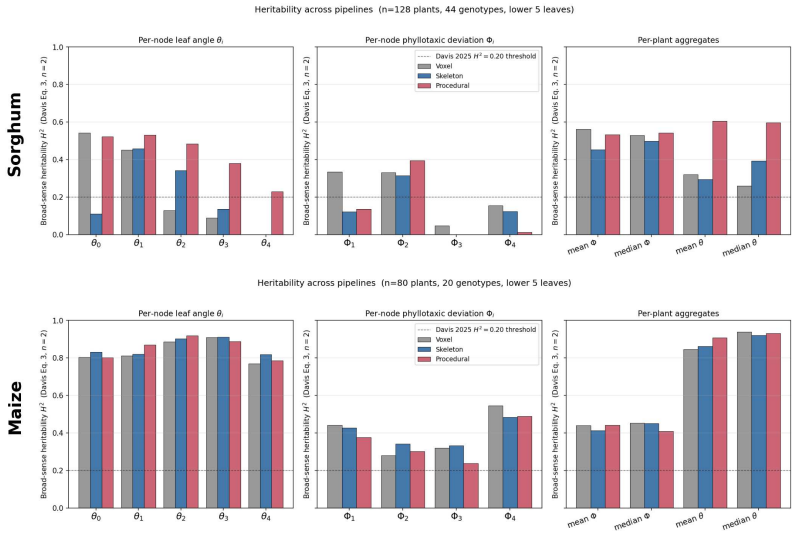


Fig. 4: Broad-sense heritability of 13 lower-canopy traits from the voxel pipeline, skeleton mode $\mathcal{S}$ and procedural mode $\mathcal{P}$. Inclination point estimates are often retained, while divergence-angle estimates can weaken.

Two trait-level contrasts illustrate the different effects of fitting. For the lowest-leaf inclination θ_0, H^2 was 0.56 in the voxel source, 0.18 in $\mathcal{S}$ and 0.58 in $\mathcal{P}$. The procedural fit uses information across each blade, whereas the base-tangent estimator depends on a few points near the collar. These are plausible measurement effects, not evidence that the fitted angle is more physically accurate.

For the highest-leaf inclination θ_4, H^2 was 0.00 in the voxel and skeleton sources and 0.35 in $\mathcal{P}$. Upper-leaf movement during rotational imaging may affect estimates near the collar [3], while the procedural fit uses the full centerline. Because fitting is independent for each leaf, this result cannot be attributed to smoothing across nodes. Divergence-angle aggregates are more sensitive to the common-pair filter and do not improve uniformly.

These results support retention of some heritable inclination signal under three-parameter fitting, with substantial uncertainty about any gain. Across $1,000$ genotype bootstrap resamples, $\mathcal{P}$ exceeded the voxel source in median H^2 in 59.2% of sorghum and 55.0% of maize samples; the 95% intervals overlapped widely. With 44 sorghum genotypes, the direction and magnitude of an overall difference remain unresolved.

On the maize data (80 plants, 20 genotypes, with repeated scans collapsed to one value per plant and trait), median H^2 was 0.87 for the voxel source, 0.91 for $\mathcal{S}$ and 0.90 for $\mathcal{P}$ (Table 3). All 13 traits passed $H^2 \geq 0.20$ in the voxel and procedural sources, and 12 did in $\mathcal{S}$.

2.6 Data-driven Generation

The PhenoFrame procedural generator (described in Section 4.2.2) samples four plant-level factors, per-node canopy curves, and leaf-level variation. It was configured using measurements from the same 324 reconstructed sorghum plants, with some shape settings adjusted. Subsequently, we generated 324 synthetic plants and measured their geometry with the trait pipeline. This process serves as an in-sample calibration check of marginal distributions rather than an independent validation of the population model.

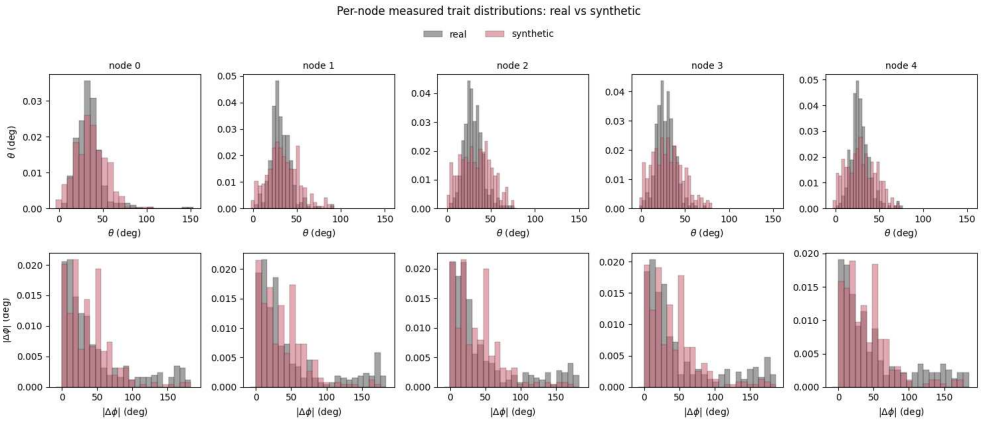


Fig. 5: Synthetic versus real leaf angle θ and divergence angle Φ distributions across the lower five nodes. Real values are published reference traits for sorghum; synthetic values are measured from generated geometry with our trait pipeline.

Across the lower five nodes, the synthetic and real leaf-inclination and divergence-angle histograms display overlapping ranges (Figure 5). Synthetic inclination exhibits a small positive bias, while divergence-angle marginals more closely track the real population. The two sets of values were measured by different pipelines, and the generator was calibrated on this population; therefore, the comparison does not establish out-of-sample realism.

The generator approximates calibrated per-node trait distributions but does not yet reproduce joint within-plant structure. The mean absolute between-node leaf-inclination correlation was 0.74 in synthetic plants compared to 0.25 in real plants, due to generated leaves sharing a plant-level angle. The current output should therefore not be used as a substitute for real within-plant correlations.

2.7 Descriptor can be used for genetic analysis

Genome-wide association analysis of $\mathcal{D}$ azimuth values from a single imaging day identified eight regions associated with the phyllotaxy of the lower four leaves (Figure 6 and Table S3). Six of these regions contain genes with plausible links to the regulation

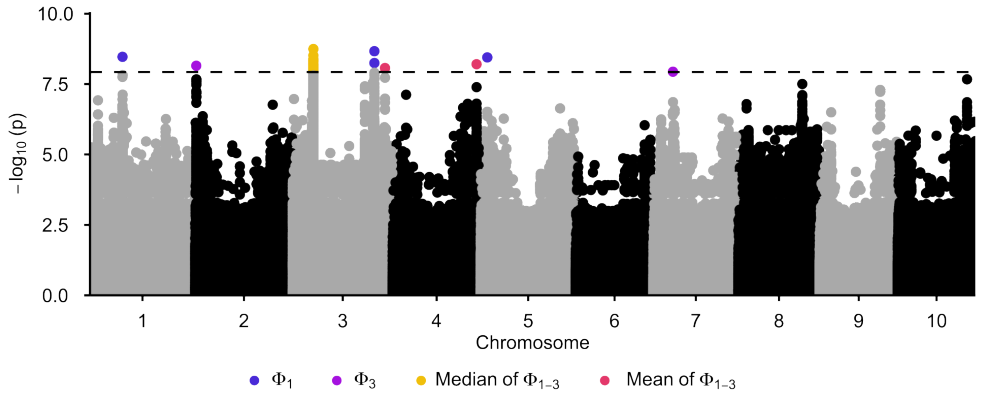


Fig. 6: **GWAS identifies genomic regions associated with $\mathcal{D}$ azimuth values.** Significant markers are colored by the trait with which they are associated, and non-significant markers are plotted in gray or black to denote which chromosome they are located on. The black horizontal dashed line at $-\log_{10}(p) = 7.92$ indicates the most conservative Bonferroni threshold across traits, based on the effective number of independent tests.

of lower canopy phyllotaxy. A region at 26.2 Mb on chromosome 1 associated with the first phyllotaxic angle contains *Sobic.001G246900*, a CUL3 E3 ubiquitin ligase substrate primarily expressed in growing internodes [39] and with both frameshift and premature stop codon variants segregating in the population. The region near 66 Mb on chromosome 3 contains both *Sobic.003G260300*, a RAX3/RGN1-like MYB transcription factor whose rice ortholog regulates axillary panicle branching by regulating the expression of the cytokinin activation enzyme, LOG [40], and *Sobic.003G259400*, a p60 subunit of katanin microtubule-severing ATPase whose orthologous loss-of-function mutants in cereals have dwarfed phenotypes with altered organ shape [41]. A second region on chromosome 3 near 74 Mb contains *Sobic.003G366000*, an OVATE family transcriptional repressor whose rice ortholog has been shown to regulate brassinosteroid signaling, leaf angle, and organ development [42]. Because auxin gradient patterns control leaf primordia in the shoot apical meristem [43], we focused on candidates acting in auxin transport and signaling. The strongest signal, at 16.7 Mb on chromosome 3, harbors an ortholog of the *Arabidopsis* trans-Golgi network membrane protein ECHIDNA, which is required for the secretion of the auxin influx carrier *AUX1* in Arabidopsis [44], in strong linkage disequilibrium with the lead marker ($R^2 = 0.94$). A region at 5.4 Mb on chromosome 5 contains Sobic.005G052500, an auxin carrier expressed in the stem whose maize ortholog imports auxin and has been shown to be drought-responsive [45]; a missense mutation in this gene is associated with a significant increase in the first phyllotaxic angle (Figure S3). The region at 67.9 Mb on chromosome 4 contains *SbER2-1*, a receptor-like kinase belonging to the ERECTA family known to regulate organ development in other species. The heterologous overexpression of this protein in maize and Arabidopsis confers increased drought resistance

and lignin deposition [46]. The identification of multiple loci with potential mechanistic links to phyllotaxy from single-plant measurements indicates that $\mathcal{D}$ azimuth values capture meaningful biological variation.

3 Discussion

PhenoFrame is introduced as a unified descriptor with two compatible modes: $\mathcal{S}$ retains observed leaf centerlines, whereas $\mathcal{P}$ replaces each with three fitted parameters. This comparison directly quantifies the trade-off: the median sorghum descriptor size decreased from 9.6 to 4.5 kB, while leaf-angle agreement with the reference decreased from $r^2 = 0.46$ to 0.36. The descriptor also supports generation and trait extraction, although these representations do not preserve all geometric or statistical properties.

This descriptor is not the first formal representation of plant architecture; its specific contributions warrant clarification. Multiscale tree graphs (MTGs) [31, 32] provide a general topological encoding of plant structure at arbitrary scales and serve as the foundation for established modeling platforms such as OpenAlea [29] and AMAPstudio [30]. MTGs are intentionally geometry-agnostic: geometry is attached to topological entities as opaque attributes, which can result in two MTGs describing the same plant carrying incompatible geometric payloads. Neither the traits derived from that geometry nor the procedures that generated it are included in the representation.

In contrast, PhenoFrame makes the opposite trade-off. It fixes a single canopy topology (alternate distichous phyllotaxy with a dominant main shoot) and, in exchange, specifies the geometry completely as a closed vector of biologically interpretable parameters with generative semantics. A descriptor instance has a defined path to mesh generation and trait measurement (Section 2.2). This path enables use as a measurement record, a generator input, and a phenotyping target; the self-consistency test quantifies the residual between declared and extracted values. The two approaches are complementary rather than competing: a PhenoFrame descriptor can be expressed as an MTG at the plant, tiller, or leaf scale, and an MTG-based platform could use these descriptors as a geometric back end for grass canopies.

The relationship to metadata standards is also complementary. MIAPPE [47, 48] specifies the required elements for recording a phenotyping experiment (study design, environment, treatments, observation variables) but remains deliberately agnostic regarding the geometric representation of individual phenotypes. PhenoFrame addresses this missing layer for grass canopy architecture, and a descriptor instance can be referenced as the value of an MIAPPE observation. In terms of established standards, MIAPPE serves as the experimental metadata envelope, while PhenoFrame fulfills the role that the OME data model [35] provides for images or SBML [36] provides for reaction networks: a shared, tool-neutral encoding of the object itself.

The GWAS demonstrates that divergence-angle traits measured from $\mathcal{P}$ can be incorporated into genetic analysis, although eight regions passed per-trait thresholds without a subsequent study-wide correction. The Φ_1 phenotype exhibited no detectable heritability in the smaller replicated panel, and several candidate-gene assignments remain unresolved. The generator reproduced marginal per-node traits but overstated between-node leaf-angle correlation. These findings represent useful

downstream demonstrations of the shared representation, but further validation and replication are required before making stronger biological claims.

A limitation is that the current schema and implementation describe only a dominant main shoot with alternate distichous leaves; other grass architectures remain untested. The procedural engine employed maize-derived blade-shape defaults for these tests.

Future work should focus on functional–structural plant models (FSPMs), as providing an FSPM compatible with this format would enable comprehensive plant simulation and testing under varying growth and environmental conditions. Integrating PhenoFrame with continuous data capture and simulation would facilitate the development of digital twins [49]. Additionally, future research should develop reconstruction algorithms that output plants in the descriptor format. While the stochastic generator can produce annotated plant geometries, it must be improved to reproduce joint within-plant trait structures before using such data to train models that depend on trait correlations. Incorporating additional reconstruction pipelines, developing independent software readers for the descriptor to further test its applicability, and evaluating the suite with raw images rather than segmented plants should be explored.

4 Methods and materials

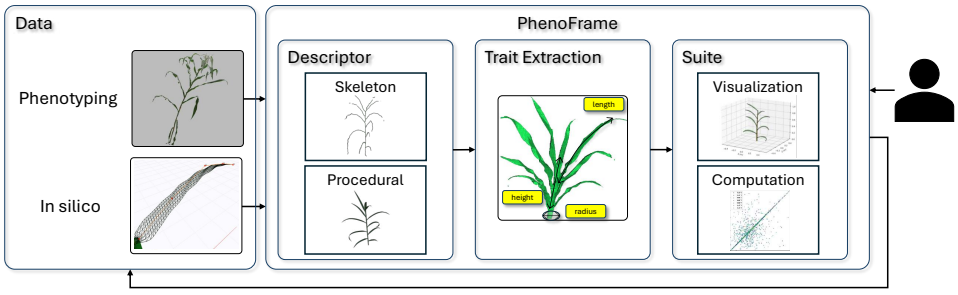


Fig. 7: PhenoFrame schema. Measured and in silico data use one descriptor with skeleton and procedural leaf modes, supporting geometry generation and trait extraction.

4.1 Plant Descriptor

The descriptor encodes a plant as a three-level hierarchy: each *Plant* contains one or more *Tillers*, each of which contains one or more *Leaves*. Each level exposes a fixed set of parameters such that any instance of the descriptor is fully specified by a vector of biologically interpretable values, together with, where required, a small set of geometric overrides. The schema is defined in the XML descriptor $\mathcal{D}$, and the suite validates every descriptor file against it before parsing.

Plant. At the plant level, a single attribute identifies the species. The descriptor schema is designed to be species-agnostic: the parameter set described below was developed for grass canopy architecture and applies to any species exhibiting alternate distichous phyllotaxy and a dominant main shoot. The reference procedural model presented here is calibrated for *Zea mays*; descriptors of other grass species (e.g., sorghum) are valid instances of the schema and can be processed by the trait pipeline. However, generating them with the maize-calibrated procedural model yields species-approximate rather than species-faithful meshes.

Tiller. At the tiller level, the descriptor specifies the geometry of a single shoot axis using geometric parameters defining stem dimensions. The current implementation supports a single tiller per plant. The schema admits multiple tillers.

Leaf. At the leaf level, the descriptor specifies the geometry and orientation of each blade. The leaf is the locus of most biologically meaningful traits and is therefore the most heavily parameterized level of the hierarchy.

4.1.1 Leaf parametrization

Each leaf (shown in Figure 8) is described by the parameters listed in the leaf level of Table S1. The geometry of a blade is constructed in three stages: (a) a center curve $\mathbf{C}(s), s \in [0,1]$, parametrized by arc length from the collar, (b) a width profile $w(s)$ that defines the blade silhouette, and (c) a pair of edge splines that are derived from $\mathbf{C}(s)$ and $w(s)$ and are controlled by leaf curl and twist. All three stages occur in a leaf-local frame, and the resulting control points feed into cubic open-uniform B-splines evaluated by de Boor's algorithm [50]. Each stage is described below.

Center curve. The center curve $\mathbf{C}(s)$ is a cubic B-spline defined by a control polyline formed by n_i consecutive points $\mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_{n_i-1}$.

All blade construction is carried out in a leaf-local frame whose z-axis points along the stem (upward) and whose y-axis is the radial direction pointing away from the stem surface, with x completing a right-handed frame; the extracted traits are subsequently re-expressed in the right-handed, Y-up world frame used for trait extraction (Section 4.3). The starting point $\mathbf{C}(0) = \mathbf{P}_0$ is placed on the stem surface, $\mathbf{P}_0 = (0, r_o, 0)$, where r_o is the stem radius at the sheath exit.

The tangent vector $\mathbf{C}'(0)$ at $\mathbf{C}(0)$ is set to $\mathbf{C}'(0) = (0, \cos\theta, \sin\theta)$ and it follows the leaf angle θ, $\mathbf{d}_{target} = (0, \cos\theta, \sin\theta)$, where 0° corresponds to a leaf horizontally exiting the stem (along the local y-axis) and 90° to a leaf exiting parallel to the stem (along the local z-axis). We model the leaf's droopiness as a quadratic rotation of the target tangent. At each step $i \in 1, \ldots, n_i - 1$, with $t_i = i/(n_i - 1)$, the target is rotated,

$$\mathbf{d}_{target}(t_i) = R_{yz}\big(a_D t_i^2\big)\mathbf{d}_{target} \tag{1}$$

where $a_D = \texttt{droopiness}$ (in degrees) and $R_{yz}(\varphi)$ is a rotation by φ in the y–z plane. This produces a tip drop similar to the one caused by gravity on the blade.

The actual stepping direction vector $\mathbf{d}_i$ relaxes toward this rotated target rather than tracking it instantaneously, which avoids kinks at the base where $\mathbf{d}_{\text{target}}$ changes rapidly:

$$\mathbf{d}_i = \text{normalize}\big((1-\lambda)\mathbf{d}_{i-1} + \lambda\mathbf{d}_{\text{target}}(t_i)\big), \qquad \lambda = \tfrac{1}{2} \tag{2}$$

with $\mathbf{d}_1 = \mathbf{d}_{\text{target}}(t_1)$ as the initial condition. Here $\mathbf{d}_i$ is a unit vector giving the local growth direction of the center curve at step i: it points from the previous control point $\mathbf{P}_{i-1}$ toward the next control point $\mathbf{P}_i$. The relaxation weight λ remains constant for the entire plant. Successive control points are placed at equal intervals $\Delta s = L/(n_i - 1)$, where $L = \texttt{leafLength}$, by:

$$\mathbf{P}_i = \mathbf{P}_{i-1} + \Delta s \cdot \mathbf{d}_i$$

The continuous center curve $\mathbf{C}(s)$ is defined as a cubic B-spline by $\mathbf{P}_i$ with an open-uniform clamped knot vector. De Boor's algorithm is used to generate 128 samples for trait extraction and 40 for visualization.

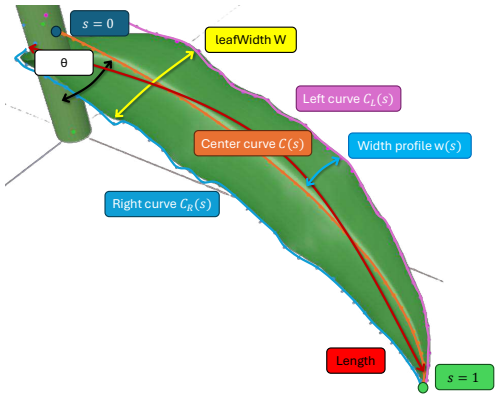


Fig. 8: Annotated parametrized leaf shows how the θ, Leaf Width W, and Length L parameters affect the leaf, as well as the Right $\mathbf{C}_R(s)$, Center $\mathbf{C}(s)$, and Left $\mathbf{C}_L(s)$ curves that make up the leaf.

The leaf **width profile** $w(s)$ defines the blade width along the center curve. It comprises two sections: a collar transition, in which the cross-section interpolates between a sheath wrapped around the stem and a flat blade, and a blade that tapers to the tip. The collar transition occupies $s \in [0, s_0]$ with $s_0 = 1/(n_i - 1)$. At $s = 0$, the left and right edge control points lie on the stem surface at azimuthal angles $\pm\alpha_c$ from the radial direction, where $\alpha_c = 135°$, so the sheath wraps about 270° around the stem. Over $s \in [0, s_0]$, $w(s)$ linearly interpolates from these wrapped positions to $\pm W/2$, where $W = \texttt{LeafWidth}$ is the maximum blade width. This produces a smooth transition from a tubular sheath to a flat blade without introducing a surface discontinuity.

For $s > s_0$, the width follows a monotonic power-law taper toward the tip

$$w(s) = W \cdot \left[1 - \left(\frac{s - s_0}{1 - s_0}\right)^{n_a}\right] \tag{3}$$

where $n_a = \texttt{widthTaper}$ controls the sharpness of the tip narrowing ($n_a < 1$ produces a sharply pointed tip, $n_a > 1$ a blunter one).

Leaf edges. Once the center spline and width profile are defined, the leaf edges are generated as left and right splines $\mathbf{C}_L(s)$ and $\mathbf{C}_R(s)$ placed at $\mathbf{C}(s) \pm \frac{1}{2}w(s)$. We also apply two modifications: leaf curl, which bends the edges inward or outward, and leaf twist, which rotates the edges around the central curve.

The descriptor also supports the explicit use of control points, enabling bypass of the procedural definition and direct use of supplied control points for a skeleton-based descriptor, denoted as $\mathcal{S}$. Only tiller-level parameters continue to apply to generate the geometry. This approach populates the descriptor from plant skeletons in Section 2.3.

The remaining parameters listed in Table S1 control aspects of geometry rendering, such as low- and high-frequency noise on the leaf edges and surface, UV texture coordinates for visual realism, and ligule and sheath geometry. These parameters do not affect blade construction or the extracted traits.

4.1.2 XML Definition

The descriptor $\mathcal{D}$ is stored as an XML document with the schema specified in Section 4.1. A minimal example is shown in Listing S1. A specified schema in XSD format [51, 52] is provided as part of the code. Each plant is a root `<plant>` element with species and phenotypeId attributes, containing one or more `<tiller>` children. Each tiller carries the stem-level architectural parameters as attributes (radius, Δr, α) and contains a `<leaves>` element whose number attribute matches the count of `<leaf>` children. Each `<leaf>` carries the architectural parameters of Table S1 as attributes. Surface-detail, ligule, and sheath parameters use schema-supplied defaults when omitted and may be specified per-leaf when needed.

All descriptors are validated against the XSD before parsing. To use the $\mathcal{S}$ representation, each child element `<leaf>` has x_n, y_n, and z_n attributes specifying each control point's coordinates in the leaf-local frame. The useCtrlOverrides attribute on `<leaf>` enables the use of $\mathcal{S}$ instead of the procedural representation.

4.2 Populating the descriptors

To evaluate the descriptor and tool suite, we populate it from two complementary data sources: real plants reconstructed from field-phenotyping images, and synthetic plants obtained from a procedural stochastic generator. Both produce valid descriptor instances that can be generated and measured with the included tool suite.

4.2.1 Voxel Skeletons

For maize, we use scans from the University of Nebraska-Lincoln Greenhouse Innovation Center comprising 80 plants (20 genotypes $\times$ 4 replicates), each imaged at 23-60 timepoints over the growth period, yielding 3,317 individual scans. We reconstruct and skeletonize each scan using the same pipeline as for the sorghum data, then convert it independently to the $\mathcal{S}$ and $\mathcal{P}$ representations. The two aggregation levels are used for different purposes: the geometric and trait accuracy analyses of Sections 2.3 and 2.4 summarize each reconstructed scan and therefore include all 3,317 repeated scans, whereas the heritability analysis of Section 2.5 requires one value per plant and

collapses each plant's timepoint series to a single per-trait value, operating on the 80 plants.

For sorghum, we use the 2018 images from the sorghum phenotyping experiment of [3, 4]. The dataset comprises 366 sorghum plants from 236 genotypes in the Sorghum Association Panel [53], imaged at the University of Nebraska-Lincoln Greenhouse Innovation Center 47 days after planting, with each plant photographed from five side views and one top view. The Zenodo archive provides voxel-reconstructed and skeletonized versions for 351 plants: 3D voxel carving of the multi-view imagery produces a 512^3 volumetric grid per plant, and a subsequent skeleton extraction stage reduces the grid to a graph classified into stem and leaf segments. Of these, 324 yielded skeletons with the topology required by our conversion pipeline (a single connected stem with leaves attached at distinct insertion heights) and were retained for the geometric and trait-accuracy experiments in Sections 2.3 and 2.4. The heritability analysis of Section 2.5 further restricts this set, following [3], by excluding the reference genotype and keeping only genotypes with at least two replicate plants, leaving 128 plants across 44 genotypes.

The descriptor $\mathcal{D}$ is populated using explicit control points for the $\mathcal{S}$ representation in Section 2.3, which preserves the skeleton geometry, and via procedural fitting of each leaf using three parameters from Table S1 for the $\mathcal{P}$ representation in Sections 2.4 and 2.5.

Skeleton representation $\mathcal{S}$. Each point $\mathbf{P}_0, \ldots, \mathbf{P}_{n_i}$ of the skeleton of each leaf can be manually specified in this representation, which then becomes the control points of the center spline $\mathbf{C}(s)$. Left $\mathbf{C_L}(s)$ and right $\mathbf{C_R}(s)$ edge splines can also be placed symmetrically about the centerline using the leaf-level `leafWidth` W supplied as a descriptor input, which will override the construction of the width profile $\mathbf{w}(s)$ with the manually input splines. The resulting descriptor bypasses the procedural model. Collar-to-collar distances are recovered as the differences in cumulative arc length along the stem skeleton between successive insertion points, and the per-leaf `leafLength` is set to the arc length of the leaf skeleton (Equation (4)).

$$L = \sum_{j=1}^{m} |\mathbf{P}_j - \mathbf{P}_{j-1}|_2 \tag{4}$$

Procedural parameter fitting. To test the procedural representation $\mathcal{P}$, `leafLength`, `leafAngle`, and `droopiness` are fit from each leaf skeleton $\mathcal{S}$. The value of `leafLength` is set directly from the skeleton arc length (Equation (4)). The remaining two parameters are estimated by nonlinear least-squares minimization of the sum of squared distances between the resampled procedural center curve and the resampled skeleton polyline, both sampled at `splinePoints` n_i. We use multiple optimizer restarts with different droopiness initializations to avoid local minima. The remaining descriptor parameters (azimuth, internode distance) are populated identically to the explicit control points path. Leaves with fewer than 20 skeleton voxels or with post-fit residual greater than the 90th percentile of the population are excluded from downstream analysis. Both conversion paths are implemented in the tool suite Python

package and produce schema-valid descriptor XML files that the procedural model and trait pipeline consume without modification.

4.2.2 In Silico Generator

The stochastic generator $\mathcal{G}$ produces descriptor instances $\mathcal{D}$ by sampling parameter values from a hierarchical model fit to a target population. The generator produces schema-valid plant instances and is calibrated to approximate the marginal trait distributions of the sorghum population analyzed in this study; we evaluated joint within-plant correlations separately.

Latent factor architecture. Each plant is defined as a descriptor $\mathcal{D}$ with 14 plant parameters (Table S1) defined. The plant structure is determined by four latent traits (Vigor, Slenderness, Posture, and Texture), each sampled per plant from an independent standard normal distribution and mapped to the descriptor parameters as summarized in Figure S4. The four traits are biologically interpretable.

Vigor. It represents the overall size of the plant; it works by positively affecting size-related parameters (stem radius, leaf count, length, width, and plant height) and negatively affecting droopiness, so that vigorous plants are larger and upright. It is the main driver of variation in plant size.

Slenderness. It represents a plant's thickness, trading it for height. This trait increases plant height and leaf length and decreases leaf width and stem radius, producing taller plants with narrower stems and blades. Because it is independent of vigor, it allows for two plants of the same overall size to differ in build, defining the plant's proportions.

Posture. This represents the orientation of the canopy, controlling how the leaves and stem are angled. It increases leaf angle, droopiness, and stem inclination, controlling stem-axis bend independent of size. It controls the part of the canopy that defines light interception, either by keeping the plant upright or allowing it to spread.

Texture. Defines color and reflectivity of the plant surface. This parameter does not affect any of the biological architecture parameters.

The generator sets plant-level values based on the population mean for each trait; the method for generating 3D meshes leverages measured trait values in conjunction with the in silico generator. For individual leaf parameters, the generator offsets the plant-level value using a canopy curve evaluated at the leaf's rank. Canopy curves are represented as a small set of control values that are interpolated to encode canopy position gradients, such as leaf angle increasing with height. Each parameter is then perturbed by Gaussian jitter with a standard deviation scaled by the set σ value, yielding a final value for each leaf. Internode distances are distributed along the plant height according to a whorl model that compresses the upper internode distances by a fraction c_w starting at a normalized stem position s_w. All elements of the generator configuration are stored in a single XML configuration file, distinct from each generated plant descriptor.

4.3 Trait Extraction

Given a descriptor, the trait pipeline reconstructs each leaf's center curve $\mathbf{C}(s)$ as defined in Section 4.1.1, evaluates it at 128 positions using de Boor's algorithm, and reports the per-leaf traits of Table S2 in a right-handed, Y-up world frame. Leaf azimuth and signed interleaf azimuth difference are circular; the divergence phenotype Φ used in heritability and GWAS is the absolute deviation of the azimuth increment from 180°.

4.4 Implementation

The PhenoFrame suite provides a Python package for descriptor parsing, serialization, validation, conversion, generation, and reference trait extraction, together with a C++17 core for mesh generation and trait extraction. Both extractors share the procedural geometry definition and agree within 0.1 mm and 0.1° for every trait in the $N = 500$ validation population. The package name and release version will be specified with the archived code.

Schema validation. A conformant XSD 1.0 schema (`descriptor.xsd`) that describes the descriptor grammar is distributed with the package; it conforms to XML Schema 1.0 [51, 52] so that any descriptor can be validated with a standard XSD 1.0 processor. Descriptor files are parsed with Python's standard library XML parser, and co-occurrence constraints that would require XML Schema 1.1 are enforced in the parser rather than by the schema.

Determinism. The stochastic generator draws all latent factors and per-leaf jitter from a single seeded pseudo-random stream, so a given configuration file and seed reproduce an identical descriptor population bit-for-bit. The procedural fitting of Section 4.2.1 uses fixed optimizer restart initializations rather than random ones, and is therefore deterministic given a skeleton. Both populations reported here use `seedstart` $= 0$, with plant i drawing seed `seedstart` $+ i$ ($N = 500$, Section 2.2; $N = 324$, Section 2.6).

Performance. Timings were measured on an AMD Ryzen 7 9800X3D (8 cores, 32 GB RAM) in a single thread. Converting one voxel skeleton to a descriptor takes $\approx$ 95 ms ($\mathcal{S}$ path) and $\approx$ 0.24 s ($\mathcal{P}$ path, including the nonlinear least-squares fit with 5 restarts); generating one plant with the stochastic generator takes $\approx$ 0.1 s; generating a mesh from a descriptor takes $\approx$ 49 ms; and extracting all traits of Table S2 from one descriptor takes a few milliseconds in either implementation ($\approx$ 5 ms C++, $\approx$ 4 ms Python reference). The full maize dataset of 3,317 scans is converted (both the $\mathcal{S}$ and $\mathcal{P}$ representations) and measured in $\approx$ 19 min on a single core. Per-plant cost is therefore about three orders of magnitude below the 10-20 min required for manual phyllotaxy measurement [3], which is what makes population-scale descriptor generation practical.

4.5 Quantitative genetic analyses

Broad-sense heritability was estimated for 13 lower-canopy traits derived from the per-leaf traits of Table S2 across 128 plants representing 44 replicated genotypes using the model presented in [3], $H^2 = \sigma_G^2/(\sigma_G^2 + \sigma_e^2/2)$ with $n = 2$ replicates, excluding the

reference genotype. Traits included the lower five leaf inclination angles ($\theta_0, \ldots, \theta_4$), the lower 4 divergence angles ($\Phi_1, \ldots, \Phi_4$), and four per-plant aggregates (the mean and median of θ and of Φ over the lower five leaves). We transformed divergence angles to their difference from the expected 180 degrees, and excluded values greater than 90 degrees before trait estimation.

Best linear unbiased estimator (BLUE) values were estimated for the median of the lower 3 divergence angles and the means of the lower 2 and lower 3 divergence angles in addition to the divergence angle traits described above. BLUE values were estimated via lm in R, modeling each trait as

$$y_i = \mu + g_i + \varepsilon_i, \tag{5}$$

where μ is the intercept and g_i is the effect of genotype i (fit as a fixed effect to obtain BLUEs).

We conducted genome-wide association studies (GWAS) for each divergence angle trait using the mixed linear model test implemented in PANICLE [14], with leave-one-chromosome-out kinship matrices and 5 principal components calculated from the genetic marker data. The genetic marker data used in GWAS comprised whole-genome resequencing markers from 925 sorghum lines aligned to the BTx623 v5 reference genome [54]. The resequencing data used to generate this marker set included a previously published Illumina dataset for the SAP panel [55] and additional whole-genome resequencing conducted via Ultima for the remaining lines. This marker set comprised $6,422,975$ markers. Marker-trait associations with p-values below 0.05, Bonferroni-adjusted by the number of effective tests estimated by PANICLE, were considered significant. Table S4 lists the number of sorghum lines, estimated effective independent tests, and significance threshold for each trait. Moderate and high-effect coding sequence variants relative to the BTx623 v5 reference genome [54] were identified using SnpEff [56], and their association with target phenotype(s) was tested via a single mixed linear model test in PANICLE, including a leave-one-chromosome-out kinship matrix and 5 principal components estimated from the genetic marker data.

Declarations

Funding. This work was supported in part by the U.S. National Science Foundation under awards No. 2412928, 2506783, and 2417510. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

This work is based upon efforts supported by the USDA-NIFA grants 2024-67021-42879 "Improving Forest Management Through Automated Measurement of Tree Geometry" and 2023-08563 "Optimizing soybean canopy architecture for efficient light and water use under climate change". The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government or NRCS. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

This work was supported by the National Science Foundation under grant IOS-2412930. JMD is supported by the National Science Foundation Graduate Research Fellowship Program, Grant No. 2034837. Zhongjie Ji was supported by the Heuermann Postdoctoral Fellowship during the development of the 20-genotype maize dataset. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. JS has equity interests in Data2Bio and Dryland Genetics and has performed paid work for Alphabet. The authors declare no other competing interests.

Ethics Declaration. Not applicable.

Generative AI use statement. The authors used AI models, including ChatGPT-5.5, ChatGPT-6.0, Claude Opus 5, Claude Sonnet 5, and Claude Fable 5.0/5.1, to assist in drafting code for data analysis and visualization and to identify spelling and grammatical errors. The authors reviewed and evaluated in detail all results produced using AI-assisted code and take full responsibility for the accuracy of the work presented in this paper.

Statistics and reproducibility. No statistical method was used to predetermine sample size. The Methods section states the structural and quality filters, and reports the divergence-angle common-pair filter with its retained fractions. The experiments were not randomized or blinded because measurements were made by deterministic software from existing reconstructions. Heritability uncertainty was assessed with 1,000 genotype bootstrap resamples using method-of-moments variance components; GWAS used per-trait effective-test thresholds without a study-wide correction.

Source data. Source data for Figures 2, 3, 4, 5 and 6 and Tables 2 and 3 at https://github.com/JorgeAskur/PhenoFrame.

Data availability. The previously published sorghum imaging, voxel-reconstruction and skeleton data are described by Gaillard et al. [4]; archive: https://zenodo.org/records/4426620. The published phyllotaxy reference traits are described by Davis et al. [3]; archive: https://github.com/jdavis-132/phyllotaxy. Maize imaging, skeletons and PCA-based reference traits: https://github.com/JorgeAskur/PhenoFrame/releases. The genetic marker set combines published SAP whole-genome resequencing data [55] with additional Ultima resequencing; the combined marker VCF file is submitted and in queue for curation and release at the time of submission and will be available from the Dryad repository at https://doi.org/10.5061/dryad.cfxpnvxpb. Descriptor instances in $\mathcal{S}$ and $\mathcal{P}$ modes, the $N = 500$ and $N = 324$ synthetic populations, generator configurations, and extracted trait tables: https://github.com/JorgeAskur/PhenoFrame.

Code availability. PhenoFrame source repository: https://github.com/JorgeAskur/PhenoFrame. Release contents: XSD schema, Python trait-extraction reference implementation, C++ procedural model and trait engine, stochastic generator, skeleton-to-descriptor conversion tools, and scripts and configurations needed to reproduce the manuscript figures and tables.

Author contributions. J.V. and B.B. conceived the descriptor and tool suite. J.V. implemented the procedural model, stochastic generator, trait-extraction pipeline, and conversion tools; performed the descriptor experiments; and analyzed those results. K.C. and Z.J. provided the maize data. J.D. and J.S. provided the sorghum imaging, reconstruction, and reference-trait data and advised on heritability analysis. S.P.M., J.D. and K.C. performed the GWAS analysis and wrote that section. B.B. and J.S. supervised the project and acquired funding. J.V. and B.B. wrote the remaining manuscript. All authors read and approved the final manuscript.

References

[1] FAO: The future of food and agriculture: Trends and challenges. Technical report, Food and Agriculture Organization of the United Nations, Rome (2017)

[2] Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A.: Yield trends are insufficient to double global crop production by 2050. PLOS ONE **8**(6), 1–8 (2013) https://doi.org/10.1371/journal.pone.0066428

[3] Davis, J.M., Gaillard, M., Tross, M.C., Shrestha, N., Ostermann, I., Grove, R.J., Li, B., Benes, B., Schnable, J.C.: 3D reconstruction enables high-throughput phenotyping and quantitative genetic analysis of phyllotaxy. Plant Phenomics **7**(1), 100023 (2025) https://doi.org/10.1016/j.plaphe.2025.100023

[4] Gaillard, M., Miao, C., Schnable, J.C., Benes, B.: Voxel carving-based 3D reconstruction of sorghum identifies genetic determinants of light interception efficiency. Plant Direct **4**, 1–16 (2020) https://doi.org/10.1002/pld3.255

[5] Tian, J., Wang, C., Chen, F., *et al.*: Maize smart-canopy architecture enhances yield at high densities. Nature **632**, 576–584 (2024) https://doi.org/10.1038/s41586-024-07669-6

[6] DeBruin, J., Aref, T., Tirado Tolosa, S., *et al.*: Breaking the field phenotyping bottleneck in maize with autonomous robots. Communications Biology **8**, 467 (2025) https://doi.org/10.1038/s42003-025-07890-7

[7] Dale, R., Oswald, S., Jalihal, A., LaPorte, M.-F., Fletcher, D.M., Hubbard, A., Shiu, S.-H., Nelson, A.D.L., Bucksch, A.: Overcoming the challenges to enhancing experimental plant biology with computational modeling. Frontiers in Plant Science **12**, 687652 (2021) https://doi.org/10.3389/fpls.2021.687652

[8] Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B., Buckler, E.S.: Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics **43**(2), 159–162 (2011) https://doi.org/10.1038/ng.746

[9] Olatoye, M.O., Marla, S.R., Hu, Z., Bouchet, S., Perumal, R., Morris, G.P.: Dissecting adaptive traits with nested association mapping: Genetic architecture of inflorescence morphology in sorghum. G3: Genes, Genomes, Genetics **10**(5), 1785–1796 (2020) https://doi.org/10.1534/g3.119.400658

[10] Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S.: TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23**(19), 2633–2635 (2007) https://doi.org/10.1093/bioinformatics/btm308

[11] Wang, J., Zhang, Z.: GAPIT version 3: Boosting power and accuracy for genomic association and prediction. Genomics, Proteomics & Bioinformatics **19**(4), 629–640 (2021) https://doi.org/10.1016/j.gpb.2021.08.005

[12] Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z.: Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLOS Genetics **12**(2), 1005767 (2016) https://doi.org/10.1371/journal.pgen.1005767

[13] Huang, M., Liu, X., Zhou, Y., Summers, R.M., Zhang, Z.: BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience **8**(2), 154 (2019) https://doi.org/10.1093/gigascience/giy154

[14] Schnable, J.C.: PANICLE: Python Algorithms for Nucleotide-phenotype Inference and Chromosome-wide Locus Evaluation. Source code: https://github.com/jschnable/PANICLE. https://pypi.org/project/panicle/

[15] Furbank, R.T., Tester, M.: Phenomics—technologies to relieve the phenotyping bottleneck. Trends in Plant Science **16**(12), 635–644 (2011) https://doi.org/10.1016/j.tplants.2011.09.005

[16] Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S.: Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics **126**(4), 867–887 (2013) https://doi.org/10.1007/s00122-013-2066-0

[17] Araus, J.L., Cairns, J.E.: Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science **19**(1), 52–61 (2014) https://doi.org/10.1016/j.tplants.2013.09.008

[18] Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J.H., Batchelor, W.D., Xiong, L., Yan, J.: Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. Molecular Plant **13**(2), 187–214 (2020) https://doi.org/10.1016/j.molp.2020.01.008

[19] Prusinkiewicz, P., Lindenmayer, A.: The Algorithmic Beauty of Plants. The Virtual Laboratory. Springer, New York (2012). https://doi.org/10.1007/978-1-4613-8476-2

[20] Allen, M., DeJong, T.: Using l-systems for modeling source–sink interactions, architecture and physiology of growing trees: the l-peach model. New phytologist **166**(3), 869–880 (2005)

[21] Cieslak, M., Khan, N., Ferraro, P., Soolanayakanahally, R., Robinson, S.J., Parkin, I., McQuillan, I., Prusinkiewicz, P.: L-system models for image-based phenomics: case studies of maize and canola. in silico Plants **4**(1), 039 (2022) https://doi.org/10.1093/insilicoplants/diab039 https://academic.oup.com/insilicoplants/article-pdf/4/1/diab039/42549198/diab039.pdf

[22] Měch, R., Prusinkiewicz, P.: Visual models of plants interacting with their environment. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 397–410 (1996)

[23] Cheng, T., Zhai, A.J., Chen, E.Z., Zhou, R., Deng, Y., Li, Z., Zhao, K., Shiu, J., Zhao, Q., Xu, Y., Wang, X., Shen, Y., Wang, S., Ainsworth, L., Guan, K., Wang, S.: Demeter: A parametric model of crop plant morphology from the real world. ICCV (2025)

[24] Zhai, A.J., Wang, X., Li, K., Jiang, Z., Zhou, J., Wang, S., Jin, Z., Guan, K., Wang, S.: CropCraft: Complete Structural Characterization of Crop Plants From Images (2026). https://arxiv.org/abs/2411.09693

[25] Wu, S., Wen, W., Xiao, B., Guo, X., Du, J., Wang, C., Wang, Y.: An accurate skeleton extraction approach from 3d point clouds of maize plants. Frontiers in Plant Science **Volume 10 - 2019** (2019) https://doi.org/10.3389/fpls.2019.00248

[26] Garrido, M., Paraforos, D.S., Reiser, D., Vázquez Arellano, M., Griepentrog, H.W., Valero, C.: 3d maize plant reconstruction based on georeferenced overlapping lidar point clouds. Remote Sensing **7**(12), 17077–17096 (2015) https://doi.org/10.3390/rs71215870

[27] Lee, J.J., Li, B., Benes, B.: Latent l-systems: Transformer-based tree generator. ACM Trans. Graph. **43**(1) (2023) https://doi.org/10.1145/3627101

[28] Boudon, F., Pradal, C., Cokelaer, T., Prusinkiewicz, P., Godin, C.: L-py: An l-system simulation framework for modeling plant architecture development based on a dynamic language. Frontiers in Plant Science **Volume 3 - 2012** (2012) https://doi.org/10.3389/fpls.2012.00076

[29] Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C., Godin, C.: Openalea:

a visual programming and component-based software platform for plant modelling. Functional Plant Biology **35**(10), 751–760 (2008) https://doi.org/10.1071/FP08084 https://connectsci.au/fp/article-pdf/35/10/751/332822/fp08084.pdf

[30] Griffon, S., de Coligny, F.: Amapstudio: An editing and simulation software suite for plants architecture modelling. Ecological Modelling **290**, 3–10 (2014) https://doi.org/10.1016/j.ecolmodel.2013.10.037 . Special Issue of the 4th International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA'12)

[31] Godin, C., Costes, E., Sinoquet, H.: A method for describing plant architecture which integrates topology and geometry. Annals of Botany **84**(3), 343–357 (1999) https://doi.org/10.1006/anbo.1999.0923 https://academic.oup.com/aob/article-pdf/84/3/343/7983752/840343.pdf

[32] Christophe Godin: Representing and encoding plant architecture: A review. Ann. For. Sci. **57**(5), 413–438 (2000) https://doi.org/10.1051/forest:2000132

[33] Okura, F.: 3d modeling and reconstruction of plants and trees: A cross-cutting review across computer graphics, vision, and plant phenotyping. Breeding Science **72**(1), 31–47 (2022) https://doi.org/10.1270/jsbbs.21074

[34] Soualiou, S., Wang, Z., Sun, W., Reffye, P., Collins, B., Louarn, G., Song, Y.: Functional–structural plant models mission in advancing crop science: Opportunities and prospects. Frontiers in Plant Science **Volume 12 - 2021** (2021) https://doi.org/10.3389/fpls.2021.747142

[35] Goldberg, I.G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P.K., Swedlow, J.R.: The Open Microscopy Environment (OME) data model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biology **6**(5), 47 (2005) https://doi.org/10.1186/gb-2005-6-5-r47

[36] Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J.: The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. Bioinformatics **19**(4), 524–531 (2003) https://doi.org/10.1093/bioinformatics/btg015

[37] Rübel, O., Tritt, A., Ly, R., Dichter, B.K., Ghosh, S., Niu, L., Baker, P., Soltesz, I., Ng, L., Svoboda, K., Frank, L., Bouchard, K.E.: The neurodata without borders ecosystem for neurophysiological data science. eLife **11**, 78362 (2022) https://doi.

org/10.7554/eLife.78362

[38] Chen, F., Stogiannidis, I., Wood, A., Bueno, D., Williams, D., Macfarlane, F., Grieve, B.D., Wells, D., Atkinson, J.A., Hawkesford, M.J., Pridmore, T., Murchie, E.H.: A conversational multi-agent AI system for automated plant phenotyping. Nature Communications (2026) https://doi.org/10.1038/s41467-026-71090-y

[39] McCormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B.D., McKinley, B., *et al.*: The sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal **93**(2), 338–354 (2018)

[40] Li, G., Xu, B., Zhang, Y., Xu, Y., Khan, N.U., Xie, J., Sun, X., Guo, H., Wu, Z., Wang, X., Zhang, H., Li, J., Xu, J., Wang, W., Zhang, Z., Li, Z.: Rgn1 controls grain number and shapes panicle architecture in rice. Plant Biotechnology Journal **20**(1), 158–167 (2022) https://doi.org/10.1111/pbi.13702 https://onlinelibrary.wiley.com/doi/pdf/10.1111/pbi.13702

[41] Martinez, S.E., Lau, K.H., Allsman, L.A., Irahola, C., Habib, C., Diaz, I.Y., Ceballos, I., Panteris, E., Bommert, P., Wright, A., et al.: Katanin promotes cell elongation and division to generate proper cell numbers in maize organs. Nature Communications (2026)

[42] Yang, C., Shen, W., He, Y., Tian, Z., Li, J.: Ovate family protein 8 positively mediates brassinosteroid signaling through interacting with the gsk3-like kinase in rice. PLoS Genetics **12**(6), 1006118 (2016)

[43] Richardson, A.E., Hake, S.: Drawing a line: grasses and boundaries. Plants **8**(1), 4 (2018)

[44] Boutté, Y., Jonsson, K., McFarlane, H.E., Johnson, E., Gendre, D., Swarup, R., Friml, J., Samuels, L., Robert, S., Bhalerao, R.P.: Echidna-mediated post-golgi trafficking of auxin carriers for differential cell elongation. Proceedings of the National Academy of Sciences **110**(40), 16259–16264 (2013) https://doi.org/10.1073/pnas.1309057110 https://www.pnas.org/doi/pdf/10.1073/pnas.1309057110

[45] Wang, B., Liu, C., Zhang, D., He, C., Zhang, J., Li, Z.: Effects of maize organ-specific drought stress response on yields from transcriptome analysis. BMC plant biology **19**(1), 335 (2019)

[46] Li, H., Han, X., Liu, X., Zhou, M., Ren, W., Zhao, B., Ju, C., Liu, Y., Zhao, J.: A leucine-rich repeat-receptor-like kinase gene SbER2-1 from sorghum (*sorghum bicolor* l.) confers drought tolerance in maize. BMC genomics **20**(1), 737 (2019) https://doi.org/10.1186/s12864-019-6143-x

[47] Papoutsoglou, E.A., Faria, D., Arend, D., Arnaud, E., Athanasiadis, I.N., Chaves, I., Coppens, F., Cornut, G., Costa, B.V., Ćwiek-Kupczyńska, H., Droesbeke, B., Finkers, R., Gruden, K., Junker, A., King, G.J., Krajewski, P., Lange, M., Laporte, M.-A., Michotey, C., Oppermann, M., Ostler, R., Poorter, H., Ramirez-Gonzalez, R., Ramšak, v., Reif, J.C., Rocca-Serra, P., Sansone, S.-A., Scholz, U., Tardieu, F., Uauy, C., Usadel, B., Visser, R.G.F., Weise, S., Kersey, P.J., Miguel, C.M., Adam-Blondon, A.-F., Pommier, C.: Enabling reusability of plant phenomic datasets with miappe 1.1. New Phytologist **227**(1), 260–273 (2020) https://doi.org/10.1111/nph.16544

[48] Ćwiek-Kupczyńska, H., Altmann, T., Arend, D., Arnaud, E., Chen, D., Cornut, G., Fiorani, F., Frohmberg, W., Junker, A., Klukas, C., Lange, M., Mazurek, C., Nafissi, A., Neveu, P., Oeveren, J., Pommier, C., Poorter, H., Rocca-Serra, P., Sansone, S.-A., Scholz, U., Schriek, M., Seren, Ü., Usadel, B., Weise, S., Kersey, P., Krajewski, P.: Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods **12**(1), 44 (2016)

[49] Pauli, D., Huppenthal, B., Wang, D., Schnable, J., Benes, B.: Digital twins for plant and crop improvement. Trends in Plant Science (2026) https://doi.org/10.1016/j.tplants.2026.06.005

[50] de Boor, C.: A Practical Guide to Splines. Applied Mathematical Sciences, vol. 27. Springer, New York (1978). https://doi.org/10.1007/978-1-4612-6333-3

[51] World Wide Web Consortium: XML Schema Part 1: Structures Second Edition. W3c recommendation, W3C (October 2004). https://www.w3.org/TR/xmlschema-1/

[52] World Wide Web Consortium: XML Schema Part 2: Datatypes Second Edition. W3c recommendation, W3C (October 2004). https://www.w3.org/TR/xmlschema-2/

[53] Casa, A.M., Pressoir, G., Brown, P.J., Mitchell, S.E., Rooney, W.L., Tuinstra, M.R., Franks, C.D., Kresovich, S.: Community resources and strategies for association mapping in sorghum. Crop Science **48**(1), 30–40 (2008) https://doi.org/10.2135/cropsci2007.02.0080

[54] Morris, G.P., Harder, A.M., Healey, A.L., McLaughlin, C.M., Rifkin, J.L., Cruet-Burgos, C., Jenkins, J.W., Shu, S., Spiekerman, J.J., VanGessel, C.J., *et al.*: A sorghum pangenome reference improves global crop trait discovery. Nature **652**(8112), 1245–1253 (2026) https://doi.org/10.1038/s41586-026-10229-9

[55] Boatwright, J.L., Sapkota, S., Jin, H., Schnable, J.C., Brenton, Z., Boyles, R., Kresovich, S.: Sorghum Association Panel whole-genome sequencing establishes cornerstone resource for dissecting genomic diversity. The Plant Journal **111**(3), 888–904 (2022)

[56] Cingolani, P., Platts, A., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M.: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *drosophila melanogaster* strain w1118; iso-2; iso-3. Fly **6**(2), 80–92 (2012)

Supplementary Information

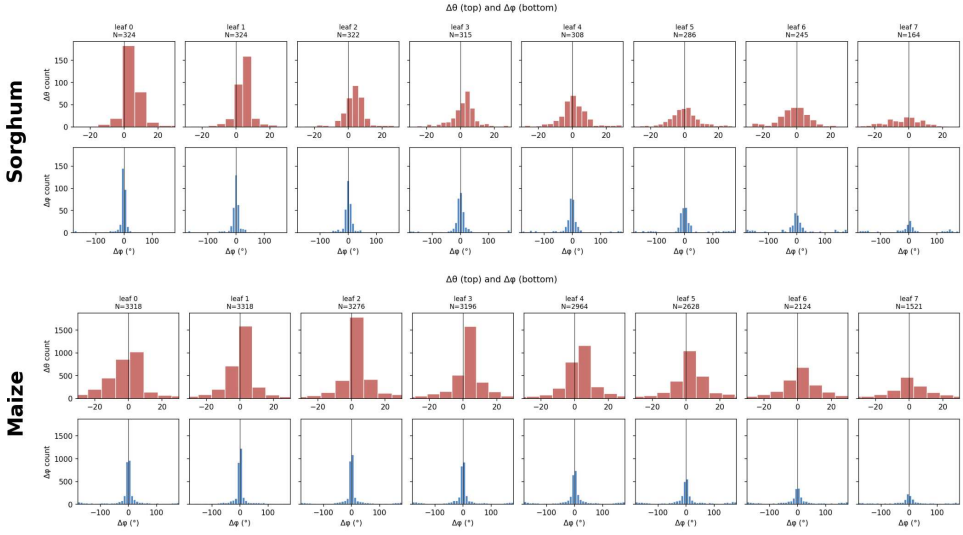


Fig. S1: Per-leaf error of skeleton mode $\mathcal{S}$ against reference leaf angle and azimuth values extracted from the same voxel skeletons (lower four leaves). Sorghum reference traits were published by Davis et al. [3]; we reimplemented the method for maize.

Level	Function	Parameter	Symbol	Definition	Unit
Plant	Identity	species	—	Species identifier	—
Plant	Identity	phenotypeId	—	Phenotype identifier	—
Tiller	Identity	type	—	Tiller type	—
Tiller	Identity	leaves/number	n_L	Number of leaves on the tiller (the `number` attribute of the `<leaves>` element)	—
Tiller	Stem geometry	radius	r_o	Stem radius at the base	m
Tiller	Stem geometry	stemShrink	Δr	Maximum radius reduction along the stem	m
Tiller	Stem geometry	alpha	α	Global tilt of the entire tiller about its base	deg
Leaf	Architecture	id	k	Leaf index along the stem	—
Leaf	Architecture	distance	d	Collar-to-collar distance from the previous leaf along the stem	m
Leaf	Architecture	leafLength	L	Blade length	m
Leaf	Architecture	leafWidth	W	Maximum full blade width; edges lie at $\pm W/2$ from the midrib.	m
Leaf	Architecture	leafAzimuthDeg	ϕ	Azimuthal orientation of the leaf around the stem axis	deg
Leaf	Architecture	leafAngle	θ	Angle between the initial midrib tangent and the horizontal direction perpendicular to the stem	deg
Leaf	Architecture	droopiness	a_D	Tip droop amplitude	deg
Leaf	Architecture	stemInclinationDeg	$\Delta\psi$	Per-leaf increment to the cumulative local stem inclination	deg
Leaf	Blade shape	widthTaper	n_a	Width taper exponent	—
Leaf	Blade shape	leafTwist	τ	Cumulative twist of the blade around its midrib	deg
Leaf	Blade shape	leafCurl	c	Lateral curl of the cross-section	—
Leaf	Blade shape	midribWidth	σ	Standard deviation of the Gaussian midrib mask applied to the UV texture; does not affect extracted traits	—
Leaf	Blade shape	midribTipTaperStart	t_t	Fraction of L at which the midrib taper toward the tip begins	—
Leaf	Surface detail	waveLAmp	—	Amplitude of the left-edge noise	m
Leaf	Surface detail	waveLFreq	—	Spatial frequency of the left-edge noise	—
Leaf	Surface detail	waveLPhase	—	Phase of the left-edge noise	rad
Leaf	Surface detail	waveRAmp	—	Amplitude of the right-edge noise	m
Leaf	Surface detail	waveRFreq	—	Spatial frequency of the right-edge noise	—
Leaf	Surface detail	waveRPhase	—	Phase of the right-edge noise	rad
Leaf	Surface detail	surfaceNoiseAmp	—	Amplitude of additive surface noise	—
Leaf	Surface detail	surfaceNoiseFreq	—	Spatial frequency of the surface noise	—
Leaf	Surface detail	midribTextureStrength	—	UV compression around the midrib	—
Leaf	Ligule & sheath	liguleWrapLengthScale	—	Scales the collar wrap length relative to the stem radius	—
Leaf	Ligule & sheath	liguleUnfoldSharpness	—	Exponent controlling the sharpness of the ligule unfolding curve	—
Leaf	Ligule & sheath	sheathOuterScale	—	Sheath outer radius as a multiple of the stem radius at the collar	—
Leaf	Discretization	splinePoints	n_i	Number of sample points along the midrib spline	—

Table S1: Description of all the available parameters in the descriptor. Parameters are categorized by function and defined.

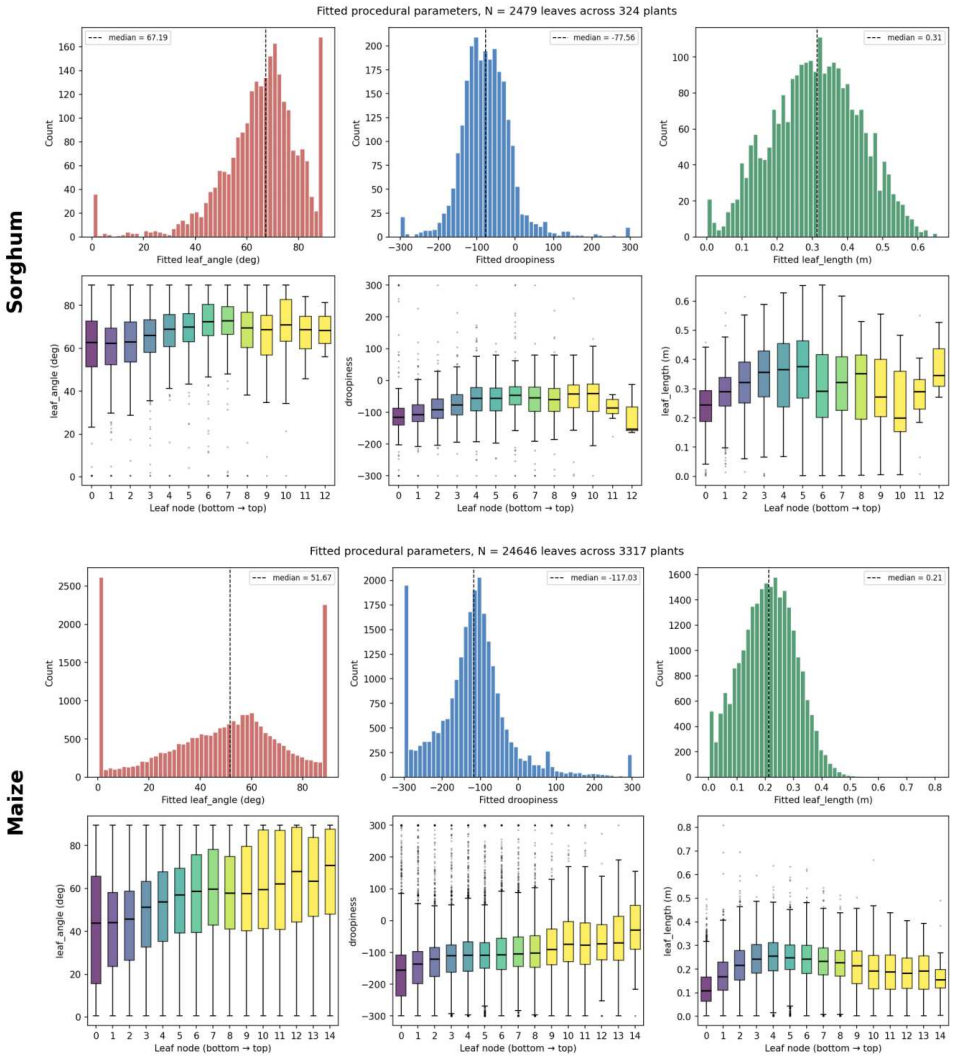


Fig. S2: Fitted procedural-mode parameters by canopy position in sorghum (324 plants) and maize (3,317 repeated scans). Trends vary with leaf rank. Some fits reach the optimizer bounds for leaf angle (0.5° and 89.5°) or droopiness (±300°); these are lower-confidence fits.

Trait	Symbol	Definition	Unit
Length	L_T	Arc length of the midrib spline	m
Leaf inclination angle	θ_t	Angle between the initial midrib tangent and the horizontal perpendicular to the stem	degrees
Leaf azimuth angle	ϕ_t	Angle in the XZ plane between the projection of the midrib base tangent and the world X-axis	degrees
Divergence-angle deviation	Φ_t	Absolute deviation of the azimuth increment from 180° between consecutive leaves	degrees
Leaf blade-stem junction (equivalent to the leaf collar for mature leaves)	$\mathbf{C}_{t0}$	Position of $\mathbf{C}(0)$ in the world frame	m (world frame)
Tip position	$\mathbf{C}_{tf}$	Position of $\mathbf{C}(1)$ in the world frame	m (world frame)
Width	W_t	Maximum blade width	m

Table S2: Extractable traits, along with their symbol, definition used for computation, and unit of measurement.

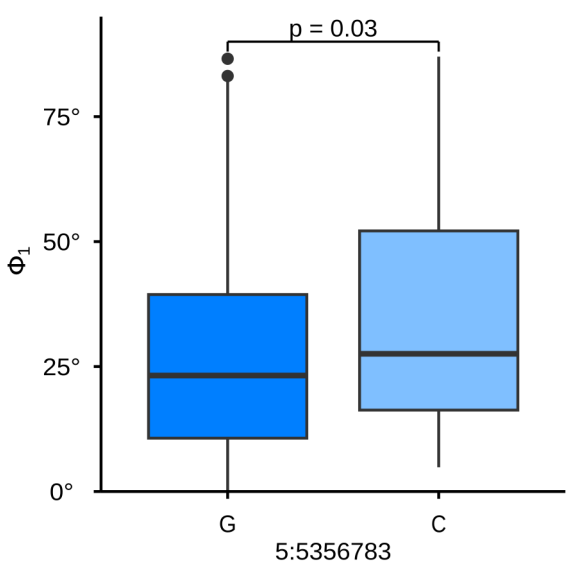


Fig. S3: **A SnpEff-annotated missense variant in the chromosome 5 region has a nominal association with higher Φ_1 ($p = 0.03$, uncorrected across tested variants).** Box plots show the 25th–75th percentiles and median; $n = 151$ (G) and 30 (C) genotypes. The candidate-gene assignment remains unresolved.

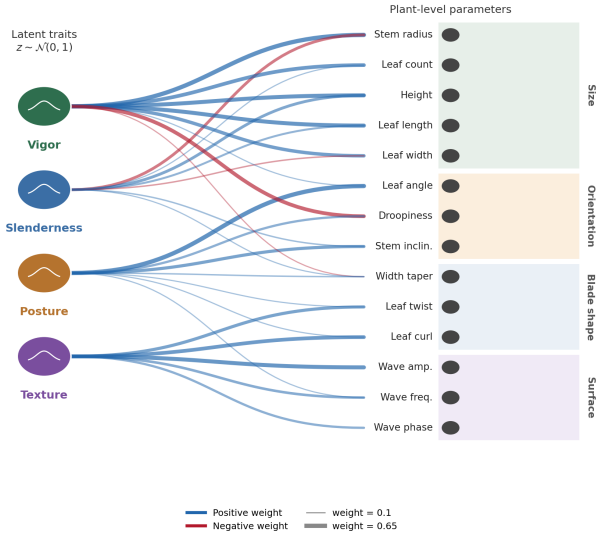


Fig. S4: Latent factor structure of the in silico generator. Each plant is generated from four latent traits (Vigor, Slenderness, Posture and Texture), sampled per plant from a standard normal distribution. Thicker edges indicate greater influence, and color indicates whether the impact is positive or negative.

Table S3: **GWAS identifies 8 regions associated with $\mathcal{D}$-derived phyllotaxy values.** Contiguous 100-kb genomic windows with 20-kb steps containing at least one significant marker were merged into a single region. The lead marker is the marker in the window most significantly associated with any trait. The peak region indicates the base-pair positions of the beginning and end of the region in megabase-pairs along the given chromosome. The trait column indicates the traits associated with the region.

Lead Marker	Chr.	Peak Region (Mb)	Traits	Candidate Gene	Putative Function
1:26324417	1	26.24 - 26.42	Φ_1	Sobic.001G246900	CUL3-based E3 ubiquitin ligase substrate adapter
2:959927	2	0.86 - 1.04	Φ_3	unresolved	unresolved
3:16659105	3	16.52 - 16.78	Mean(Φ_{1-3}), Median(Φ_{1-3})	Sobic.003G153700	TGN-localized regulator of secretory trafficking
3:66056721	3	65.96 - 66.16	Φ_1	Sobic.003G260300, Sobic.003G259400	
3:74651437	3	74.56 - 74.74	Mean(Φ_{1-3})	Sobic.003G366000	OVATE family transcriptional repressor influencing growth
4:67905663	4	67.82 - 68.00	Mean(Φ_{1-3})	Sobic.004G317200	Receptor-like kinase enhancing drought tolerance and water-use efficiency
5:5408361	5	5.32 - 5.50	Φ_1	Sobic.005G052500	Auxin influx carrier influencing root and flower development
7:15945992	7	15.86 - 16.04	Φ_3	unresolved	

Trait	Genotypes	Effective Independent Tests	Significance Threshold
Φ_1	181	4,200,006	1.19e-8
Φ_2	182	4,226,942	1.18e-8
Φ_3	167	4,215,341	1.18e-8
Φ_4	170	4,192,509	1.19e-8
Mean(Φ_{1-2})	205	4,224,725	1.18e-8
Mean(Φ_{1-3})	211	4,222,695	1.18e-8
Mean(Φ_{1-4})	212	4,222,642	1.18e-8
Median(Φ_{1-3})	211	4,222,695	1.18e-8
Median(Φ_{1-4})	212	4,222,642	1.18e-8

Table S4: Number of genotypes for which both phenotypic and genotypic data passed filtering criteria, effective independent tests estimated by PANICLE, and p-value below which associations were considered significant for each trait.

Listing S1 Minimal architecture-only descriptor for a two-leaf plant. Lengths are in meters and angles in degrees.

```
<?xml version="1.0" encoding="UTF-8"?>
<plant species="Sorghum bicolor" phenotypeId="BTx623">
  <tiller type="main"
          radius="0.0123"
          alpha="0"
          stemShrink="0.0080">
    <leaves number="2">
      <leaf id="0"
            distance="0.0815"
            leafLength="0.438"
            leafWidth="0.0438"
            leafAzimuthDeg="4.84"
            leafAngle="12.71"
            droopiness="-15.37"
            stemInclinationDeg="-0.40" />
      <leaf id="1"
            distance="0.0815"
            leafLength="0.452"
            leafWidth="0.0652"
            leafAzimuthDeg="182.13"
            leafAngle="24.23"
            droopiness="-17.14"
            stemInclinationDeg="-0.01" />
    </leaves>
  </tiller>
</plant>
```