Received: 10 April 2024 | Revised: 15 January 2025 | Accepted: 17 January 2025

DOI: 10.1002/ppp3.10644

Plants People Planet PPF Open Access

RESEARCH ARTICLE

Characterizing season-long floral trajectories in cotton with low-altitude remote sensing and deep learning

Jeevan Adhikari[1,2] | Daniel Petti[3] | Deepak Vitrakoti[1] | Wiriyanat Ployaram[1] | Changying Li[3] | Andrew H. Paterson[1]

[1]Plant Genome Mapping Laboratory, The University of Georgia, Athens, Georgia, USA

[2]CoverCress Inc, St Louis, Missouri, USA

[3]Department of Agricultural and Biological Engineering, The University of Florida, Gainesville, Florida, USA

Correspondence
Andrew H. Paterson, Plant Genome Mapping Laboratory, The University of Georgia, Athens, GA, USA, 30602.
Email: paterson@uga.edu

Changying Li, Department of Agricultural and Biological Engineering, The University of Florida, Gainesville, FL, USA, 32611.
Email: cli2@ufl.edu

Funding information
This work was supported by the National Science Foundation Plant Genome Research Program (DBI-0817707) and Growing Convergence Research (Award No. 1934481); USDA National Institute of Food and Agriculture (Award No. 2023-67021-40646); Cotton Incorporated; and Georgia Cotton Commission.

Societal Impact Statement

Plant breeding is a critical tool for increasing the productivity, climate resilience, and sustainability of agriculture, but current phenotyping methods are a bottleneck due to the amount of human labor involved. Here, we demonstrate high-throughput phenotyping with an unmanned aerial vehicle (UAV) to analyze the season-long flowering pattern in cotton, subsequently mapping relevant genetic factors underpinning the trait. Season-long flowering is a complex trait, with implications for adaptation of perennials to specific environments. We believe our approach can improve the speed and efficacy of breeding for a variety of woody perennials.

Summary

- Many perennial plants make important contributions to agroeconomies and agroecosystems but have complex architecture and/or long flowering duration that hinders measurement and selection. Iteratively tracking productivity over a long flowering/fruiting season may permit the identification of genetic factors conferring different reproductive strategies that might be successful in different environments, ranging from rapid early maturation that avoids stresses, to late maturation that utilizes the full seasonal duration to maximize productivity.
- In cotton, a perennial plant that is generally cultivated as an annual crop, we apply aerial imagery and deep learning methods to novel and stable genetic stocks, identifying genetic factors influencing the duration and rate of fruiting.
- Our phenotyping method was able to identify 24 QTLs that affect flowering behavior in cotton. A total of five of these corresponded to previously identified QTLs from other studies.
- While these factors may have different relationships with crop productivity and quality in different environments, their determination adds potentially important information to breeding decisions. With transfer learning of the deep learning models, this approach could be applied widely, potentially improving gains from

Jeevan Adhikari and Daniel Petti contributed equally to this work.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.

selection in diverse perennial shrubs and trees essential to sustainable agricultural intensification.

KEYWORDS
cotton, deep learning, flowering time, genome mapping, high-throughput phenotyping, remote sensing, unmanned aerial vehicles

1 | INTRODUCTION

Understanding the genetic controls underlying the adaptability of perennial plants to local environments has been a long-standing need in plant biology. The initiation of floral transition (switch from vegetative to reproductive growth) and its duration determine the reproductive success of plants and adaptation to their local environment. Genetic control of the floral transition and flowering patterns in perennial woody plants is complex and has yet to be adequately investigated due to difficulties in genetic modification and long vegetative growth periods (GPs) before floral initiation. While the molecular mechanisms controlling floral regulation has been extensively investigated in model annual and herbaceous crops like Arabidopsis and rice (Fankhauser & Chory, 1997; Leal Valentim et al., 2015; Wang et al., 2014), the regulation of flowering in perennial woody plants has just begun to be elucidated (Grover et al., 2015). A rather longer life cycle of these plants and an extravagant waiting time for the first floral initiation accompanied by long and variable flowering duration not only pose challenges in investigating these phenomena but also hinder timely generation of populations suitable for such studies.

Cotton (*Gossypium*), a genus of woody perennials that are widely cultivated as annuals, exemplifies the challenges associated with deciphering the genetic controls underlying the floral transition and flowering patterns of perennials. Having evolved as a woody perennial, cotton has arguably the most complex vegetative and reproductive morphology of any annual crop grown. Its growth habit produces both vegetative and reproductive organs simultaneously. The sympodial flowering pattern of the cotton plant causes a very complex production-distribution pattern of carbohydrate throughout the structure of the crop. In addition, this pattern of floral initiation, which varies by species and accessions, creates complexity and ambiguity in quantification of this very important plant characteristic.

Flowering habit in cotton differs greatly between the two predominant cultivated species, *G. hirsutum* L. (Upland Cotton) and *Gossypium barbadense* L. (Pima or Egyptian cotton), allowing for investigation of this important phenomenon in crosses made from elite domesticated cultivars belonging to these two mostly grown species. Considering the complex nature of flowering response in cotton (Kohel et al., 1974; Lewis & Richmond, 1960; Waddle et al., 1961) and limited knowledge of the flowering mechanism at the molecular level, identification of genetic signatures governing these traits would provide insight into this vital biological phenomenon.

Flowering in perennials can be spread over months, and interspecific experimental cotton populations likewise have nonsynchronous flowering, making it a complex phenotype to measure. Different variants have been used to characterize flowering in cotton such as "day of first flower," "days to 50% flowering," "flower to bud time," "nodes of first fruiting branch" (Grover et al., 2015; Guo et al., 2008; Kushanov et al., 2021; Li et al., 2013; Zhang, Jia, et al., 2021). Such variations in defining, naming, and collecting data related to these traits add to the complex nature of the trait itself. High-throughput phenotyping (HTP) systems minimize ambiguity and irregularities in extracting flowering-related data (Jiang & Li, 2020), making it feasible to count the number of flowers from images of experimental plots over the entire growing season with greater accuracy and higher spatial and temporal resolution than visual scoring methods.

While ground-based HTP systems have been routinely used to collect flowering-related data from cotton (Jiang & Li, 2020; Tan et al., 2023), confluence of plants at later periods in the growing season makes these systems less efficient. Unmanned aerial systems (UAS) are advantageous over ground-based systems in that they can be used at any stage of crop growth, and they also cover more ground in less time than ground-based systems, thus reducing the effects of solar incidence angles and other temporal factors affecting the quality of data collection (Herr, 2023). While unmanned aerial vehicle (UAV) systems have a slight disadvantage of not capturing flowers embedded deep inside the crop canopy thus curtailing the extraction of exact flower counts, the data collected from this system is adequate to get the pattern/trend on flowering behavior in cotton (Xu et al., 2018). With these data, we can investigate how genotypes differ in terms of flowering imitation, duration, cessation, and peak flowering times.

The general problem of automatically counting plant organs in images has been studied intensively for at least two decades, and a variety of techniques have been developed. The earliest approaches were generally simplistic and relied on color thresholding (Adamsen, 2000). Though some modern work still employs this technique (Thorp, 2016; Wei et al., 2018), much recent attention has been focused on Convolutional Neural Networks (CNNs) due to their proven performance on other object-counting tasks (Cohen, 2017; Li et al., 2021; Marsden et al., 2018). For instance, a customized CNN model was proposed to directly regress the number of maize tassels in its input image (Lu et al., 2017). This approach has been successfully adapted to wheat spikes as well (Madec, 2019; Xiong, 2019). Other custom frameworks have been developed in order to take advantage of simpler, less costly data annotations (Akiva, 2020; Bellocchio, 2019; Giuffrida, 2019; Petti & Li, 2022). In recent years, single-stage detectors, such as YOLO (Redmon, 2016), have gained popularity due to accuracy and efficiency. For instance, cotton seedling counting was addressed with a YOLOv4 model and a multi-object tracking approach from video frames (Tan et al., 2022), resulting in faster and slightly more accurate detections than two-stage object-detection models (Jiang et al.,

2020). More recently, an improved YOLOv8 detector has been proposed (Terven et al., 2023), providing an additional performance boost for most tasks.

There are relatively few studies that combine HTP with genome-wide association studies (GWAS). Most of them use fairly simplistic HTP methods, relying on a large number of hand-crafted features that are easy to extract from raw sensor data (Li et al., 2020; Loladze et al., 2024; Wu et al., 2021; Yang et al., 2014). Some studies gravitate towards more traditional features such as plant height (Adak et al., 2021; Spindel et al., 2018; Wang et al., 2019) or vegetation fraction (Ogawa et al., 2020). These traits are easy to extract from UAV data. Spindel et al. (2018) was even able to estimate Leaf Area Index (LAI), total fresh biomass, and vegetative biomass using a commercial UAV system and software package.

While both natural populations and designed crosses are amenable to genetic analysis, near-isogenic line (NIL) populations offer an advantage of eliminating most of the background noise due to chromosomal segments segregating in early generation populations, thus increasing the precision with which one can map genomic regions controlling these traits. Here, we used high-throughput UAS to collect flower data over the entire growing season for 820 reciprocal NILs, testing single chromosomal segments that collectively cover 71.48% and 78.72% of the *G. hirsutum* "Acala Maxxa" and *G. barbadense* "Pima S6" genomes, respectively. By applying state-of-art deep learning algorithms to count flowers, we extract and study season-long floral trajectories, identifying genetic differences that might be employed in selective breeding. With small changes in the counting algorithm to recognize different flowers/fruits, this approach could be applied widely, potentially improving gains from selection in diverse perennial shrubs and trees important to sustainable agricultural intensification.

2 | MATERIALS AND METHODS

2.1 | Population development

Plant materials used in this study were developed from a set of reciprocal crosses between *Gossypium hirsutum* L. acc. Acala Maxxa and *G. barbadense* L. acc. Pima S6 (Acala Maxxa (♀) × Pima S6 (♂)—hereafter referred to as Acala Maxxa background; and Pima S6 (♀) × Acala Maxxa (♂)—hereafter referred to as Pima S6 background), backcrossing them to respective recurrent parents for five generations using a single-seed descent approach. The details of population advancement until BC5F1 and their genotyping at BC5F1 generation have been described in Adhikari et al.. After five generations of backcrossing, 179 BC_5F_1 plants from the Acala Maxxa background and 190 BC_5F_1 plants from the Pima S6 were self-pollinated and a total of 8364 BC_5F_2 plants (2–32 individuals in each BC_5F_2 family) were grown at Iron Horse Farm, Watkinsville, Georgia, in 2019 under cultural conditions consistent with commercial irrigated cotton production. Individual BC_5F_2 plants found to contain only one introgressed segment from the donor parent were deemed NILs and were selfed to increase seeds for phenotypic evaluation. We identified a total of 397 NILs in the Acala Maxxa background and a total of 423 NILs in the Pima S6 background. These NILs (BC5F3) were grown at Iron Horse Farm, Watkinsville, Georgia, in 2021 under cultural conditions consistent with commercial irrigated cotton production.

2.2 | Genotyping

BC_5F_1 were genotyped using genotyping by sequencing (see Adhikari et al., 2023 for details). The genomic compositions of BC_5F_2 plants were inferred based on targeted microsatellite (simple sequence repeat [SSR]) genotyping of the introgressed chromosomal segments identified in their respective BC_5F_1 parents. At least two (and at most four) SSR markers were used to verify most of the introgressed regions while for small introgressions only one available SSR marker was deployed. A total of 852 polymorphic SSR markers spanning the introgressed regions were derived from several published genetic maps of crosses between *G. barbadense* and *G. hirsutum* stored in the CottonGen SSR database (https://www.cottongen.org/data/download/marker). A total of 47 published SSRs were found to be monomorphic in pilot studies and were discarded. Another 23 with ambiguous bands were also discarded, and a total of 782 SSR markers were used to genotype the BC_5F_2 progenies. Among the 8364 BC_5F_2 individuals planted in 2019, a total of 5315 (with corresponding BC_5F_1 parents carrying two to five introgressions), plants were SSR genotyped for the presence (or absence) and nature (homozygous vs. heterozygous) of the respective introgression/s that their BC_5F_1 parent carried. Individual BC_5F_2 plants that were verified by SSR markers to carry only one introgression (homozygous or heterozygous) from the donor parent were deemed as NILs.

2.3 | Field experimentation

These NILs were grown at Iron Horse Farm, Watkinsville, Georgia, in 2021 using a completely randomized design (CRD) with two replications of each NIL. Each genotype was planted in 1 m long plots, and 10 seeds were dropped in each plot using a two-row seed planter. The field was uniform and was irrigated with an overhead sprinkler system. Standard cultural practices recommended for the crop and the region were performed at recommended growth stage of the crop.

2.4 | Data collection

Data were collected twice a week from August 9, 2021, through November 5, 2021. The decision to scan the field with the UAV system during this time frame was based on historical patterns of floral initiation in these two cultivars at the Watkinsville, GA, location from the time of planting and from frequent field visits in the experimental year. Some sessions were skipped due to inclement weather conditions, resulting in a total of 23 sessions. Data collection was halted

after the first overnight freeze, after which most of the plants showed a significant drop in the number of flowers produced. Images of the field were collected using a Matrice 100 drone (DJI, Shenzhen, China) fitted with a custom mount and equipped with a Lumix G7 camera (Panasonic Corporation of North America, Newark, N.J., USA) and a 17 mm lens. The drone was flown at a height of 15 m, resulting in a GSD of 0.23 cm/px. In a few cases, technical issues with the Matrice 100 data required the substitution of equivalent data from a DJI Phantom 4 Pro v2 drone. The images were geo-referenced using a total of six ground control points distributed throughout the field, with their exact positions measured using a Real-time-Kinematic (RTK) GPS.

2.5 | Dataset composition

The dataset used for training the counting model is an expanded version of the UGA2020V dataset from Tan et al. (2023). The original ground imagery in that dataset was supplemented with a random sampling of the aerial imagery described previously. The testing and validation splits consist of aerial images randomly selected from supplementary data collected during the 2016 and 2018 growing seasons using a similar method. This composition was intended to ensure a validation process that could fairly gauge the model's ability to generalize, specifically by not including data from the same year in the training and test/valid splits. As such, the test and validation splits should, if anything, err on the side of underestimating the model's performance on the 2021 data. Table 1 provides an overview of the dataset composition.

All of the images used in the dataset are of relatively high resolution, allowing individual flowers to be distinguished. There is significant variation, however, in plant phenotypes, growth stages, and lighting conditions (Figure 1). We also note that, due to uneven germination rates across the experimental field, there is significant variation in the number of cotton plants in each plot. (This is somewhat visible in Figure 2b.) Though the completely randomized and

TABLE 1 Parameters for the imagery sources used in this dataset.

Source	Split	Img. Size (px)	No. of images	Year
UGA2020V	Train	1920 × 1080	1486	2020
M100 + Lumix G7	Train	1148 × 862	976	2021
M100 + Lumix G6	Validation	1152 × 864	160	2016
M100 + Lumix G6	Test	1152 × 864	320	2018

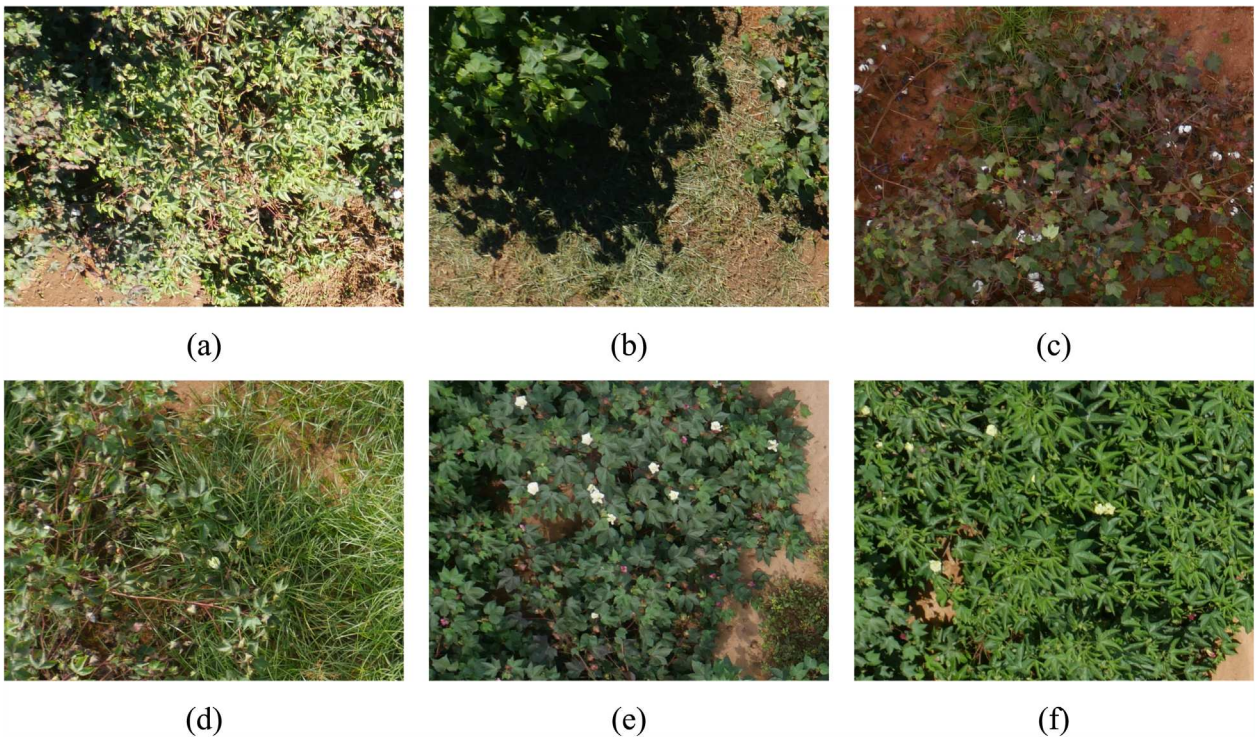

FIGURE 1 Example images from the cotton blossom dataset, showing various challenging conditions. These include poor lighting (a) and extreme shadows (b), variation in growth stages, including the presence of open bolls (c), intense weed pressure (d), and variations in phenotypes, encompassing Acala Maxxa-like (e) and Pima S6-like (f) cultivars.

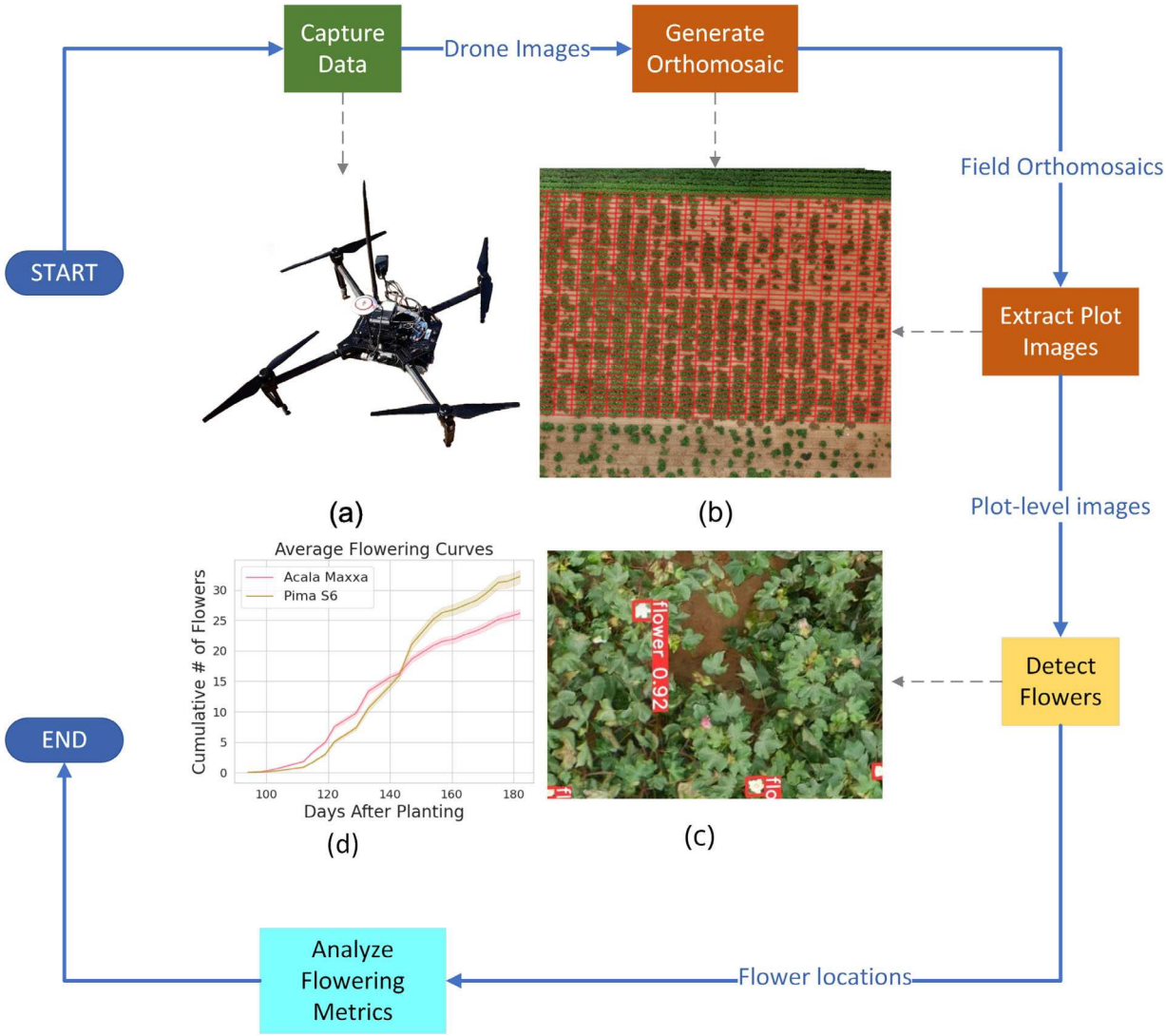

FIGURE 2 Image data collection and processing workflow. Images of the field were captured using a UAV (a). The images were stitched into an orthomosaic, and individual plots were extracted (b). The flower detection model was applied to the plot images (c), and these counts were used to extract the traits used for genetic analysis (d). The shaded regions of (d) represent a 95% confidence interval.

replicated design of our study should mitigate these environmental effects, we note that these density variations further increase the difficulty of training a flower detection model that performs reliably. Overall, the dataset presents a challenging object-detection task.

2.6 | Image analysis and phenotype extraction

Orthomosaic images were generated from each flight using Agisoft Metashape (Agisoft LLC, St. Petersburg, Russia). Plot boundaries were specified manually using QGIS (QGIS Development Team, 2009), and plot images were extracted for all orthomosaics using these boundaries. For each session, 2150 individual plot images were extracted, resulting in a total of 41,696 images for the entire season. Flower counting was performed using a YOLOv8-medium detector (Terven et al., 2023) trained on a custom dataset of flower images. This detector seemed to provide a good balance between accuracy and speed (it had to be applied to tens of thousands of high-resolution plot images during the analysis process).

Ghosal et al. (2019) proposed a pseudo-active-learning approach to training detector models, which was adopted here to expand the dataset. The model was first bootstrapped on the "UGA2020V" dataset from Tan et al. (2023), which consists of images of flowering cotton plants collected using a tractor-mounted camera. Afterwards, 10 of the original, uncropped images from the aforementioned drone data were selected at random. The model was applied to these images in order to produce initial flower detections, which were then corrected by humans using the CVAT tool (Sekachev et al., 2019). Finally, the model was retrained with the new annotations, and the process was repeated with 10 new random images, until the model

performance saturated. In this way, a large, annotated dataset was constructed with minimal human effort.

Following this analysis process, automated flower counts were extracted for every plot image. From these count data, five flowering traits were extracted as follows:

1. **Flowering start time:** Defined as the start of flowering and estimated by the time at which 5% of the total cumulative flowers have been counted, measured in days after planting (DAP).
2. **Flowering end-time:** Defined as the cessation of flowering and estimated by the time at which 95% of the total cumulative flowers have been counted, measured in DAP.
3. **Flowering peak time:** Defined as the floral flush time when each genotype produced the highest number of flowers and estimated by the session with most flower counts, measured in DAP.
4. **Flowering duration:** Defined as the difference between the flowering end-time and the flowering start time, measured in days.
5. **Inflection period slope:** Defined as the rate of change of flowers produced each day during the inflection period, measured in flowers/day. The inflection period was defined as the period between the flowering start and peak times. The slope was estimated using simple linear regression.

The entire analysis process was automated using Kedro (Bălan et al., 2020) and required around an hour to produce flower count estimates for every plot when running on an NVIDIA A100 GPU. Figure 2 provides an overview of the analysis process. The pipeline takes as input the plot images extracted from QGIS, and first extracts flower counts from each plot image. It then computes the values of the flowering traits and outputs a genotype-level analysis of flowering habit.

2.7 | Phenotypic evaluation

Pearson's correlation analysis was performed in R to look at the pairwise covariance of the traits. Phenotypic analysis of the reciprocal populations was conducted in R software, by single marker analyses using the "scanone()" function in the R/qtl package (Broman & Sen, 2009). The significance threshold was set to LOD of three, to mitigate the multiple-comparison problem. Phenotypic variance explained by each locus was reported by taking the most significant marker as independent variable and phenotypic value as dependent variable. Additive effects were estimated by half the difference of phenotypic values between the lines carrying the homozygous introgression and lines not carrying the introgression. Dominance effects were estimated by the difference of phenotypic values between the lines carrying the heterozygous introgression and the remaining lines that do not carry the introgression. If multiple or overlapping introgressions were present at both homozygous and heterozygous state, the estimation of additive effects utilized the lines carrying the introgression at homozygous state only, and the estimation of dominance utilized the lines carrying the introgression at heterozygous state only.

Gene actions for the QTLs were determined by calculating the degree of dominance (absolute values) for each QTL. The degree of dominance is the ratio of dominance effect to additive effect (d/a) of the QTL and based on this ratio; gene action of the QTLs can be categorized as (i) additive ($0 < d/a < 0.2$) (ii) partially dominant ($0.2 < d/a < 0.8$) (iii) dominant ($0.8 < d/a < 1.2$), and (iv) overdominant ($d/a > 1.2$). QTLs with dominant and overdominant effects are considered to have heterotic effects. For each QTL identified in the study, the genomic region spanning 50 kb on each side of the most significantly associated marker was used for in silico analysis. The DNA sequence from this tightly linked region was used to look for *G. hirsutum* genes in the CottonGen database, and these genes were then analyzed for biological functions, with particular focus on flowering related functions and expression on floral organs.

We also looked at previous literature to see if the QTLs identified in our study mapped to genomic locations identified by previous similar studies. However, due to vast differences in approaches to defining, capturing, and analyzing the phenotypes, the comparison would be that of apples to oranges. Instead, we utilized a unique method of identifying if QTLs in our study were in nonrandom correspondence with those reported in previous studies. For this, we used a method of aligning genetic maps and QTLs that uses the hypergeometric probability function (Feltus et al., 2006). QTLs were assumed to be orthologous (co-occurring) if they explained a significant proportion of variation for a directly comparable trait measured and the 1-LOD confidence intervals overlapped. The hypergeometric probability function equation is as follows:

$$p = \frac{\binom{l}{m}\binom{n-1}{s-m}}{\binom{n}{s}}$$

where, p is the probability of nonrandom correspondence of QTLs being compared for a given trait; n is the number of intervals, which can be compared (each interval is usually defined as approximating a QTL likelihood interval, i.e., 30 cM) along the entire genome; m is the number of matches declared between QTLs; l is the total number of QTLs identified in the larger sample; and s is the number of QTLs identified in the smaller sample.

3 | RESULTS AND DISCUSSION

3.1 | Genomic structure of reciprocal NILs

The NIL population in the Acala Maxxa background consisted of 397 individuals with only one chromosomal segment introgressed from the donor parent (*Gossypium barbadense* "Pima S6"), while that in the Pima S6 background consisted of 423 individuals (Adhikari et al., 2023). In total, these lines covered 78.72% and 71.48% of the donor genome, respectively, in the Acala Maxxa and Pima S6 backgrounds (Table 2, Figure 3, Figure 4). Single introgressed segments were identified for all chromosomes except chromosome 24 in the Acala Maxxa background and for all chromosomes except

TABLE 2 Donor genome covered by reciprocal near-isogenic line (NIL) populations.

	Acala Maxxa background		Pima S6 background	
Chromosome	No. of NILs	Coverage (%)	No. of NILs	Coverage (%)
Chr 01	51	97.14	24	91.09
Chr 02	27	95.53	19	94.08
Chr 03	9	88.58	22	87.66
Chr 04	17	85.93	16	92.44
Chr 05	30	92.67	36	82.66
Chr 06	2	1.33	34	94.59
Chr 07	20	90.01	25	87.79
Chr 08	15	93.61	41	82.72
Chr 09	26	91.76	16	89.52
Chr 10	15	85.33	18	51.15
Chr 11	14	82.49	29	63.76
Chr 12	26	85.74	7	74.46
Chr 13	17	85.68	45	90.34
Chr 14	2	92.19	11	73.13
Chr 15	14	86.35	15	74.88
Chr 16	13	89.38	10	57.51
Chr 17	11	87.79	1	70.20
Chr 18	6	83.70	15	80.20
Chr 19	3	84.08	5	86.26
Chr 20	10	91.75	1	16.71
Chr 21	31	64.55	5	72.92
Chr 22	7	76.74	6	42.33
Chr 23	12	83.20	5	64.41
Chr 24	0	0.00	10	66.88
Chr 25	10	39.18	0	0.00
Chr 26	9	91.85	7	71.01
Total	397	78.71	423	71.49

chromosome 25 in the Pima S6 background. Individual lines contained an average of 0.904% of the donor genome, ranging from 0.07% (1.37 Mb) to 2.92% (56.65 Mb) in the Acala Maxxa background while in the Pima S6 background, individual lines contained an average of 0.97% of the donor genome, ranging from 0.15% to 3.07%.

3.2 | Performance of the flower counting model

Overall, the flower counting model achieves a mean Average Precision (mAP)@0.5 of 0.87 on the validation dataset (Table 3). The mAP@0.5:0.95 is 0.54. The model was trained for 160 epochs, after which we stopped because the testing loss was no longer decreasing (Figure 5). The model performed under heavy occlusion and under variable lighting conditions, only detecting flowers on the single day that they are unpollinated (after which the petals turn pink, Figure 6d). Although seldom as accurate as ground-based approaches (Jiang et al., 2020; Tan et al., 2023), the obvious advantage of flower counting via aerial imagery is throughput. Jiang et al. (2020), though collecting data at a similar cadence using proximal sensing, only managed to image 116 individual plants.

Additionally, we found that it is clearly possible to perform useful trajectory analysis even with partial counts only. Implicit here is the assumption that systematic errors in our counting method, such as those caused by missing flowers in the earlier part of the season that are excluded by the canopy remain constant across genotypes. In this study, we found no specific reason to challenge this assumption, but it remains a consideration for any future work. Generally speaking, the flowering metrics that we derived were suitable for elucidating genotype-level variation in flowering pattern. Even so, it should be noted that the nature of some metrics, such as flowering slope, is dependent on the number of plants in each plot; therefore, care must be taken when comparing flowering metrics across different experiments.

The "pseudo-active learning" procedure described by Ghosal et al. (2019) was effective as a method for expanding the dataset size and improving the performance of the counting model (Table 3). Surprisingly, even the baseline model trained only on ground imagery was

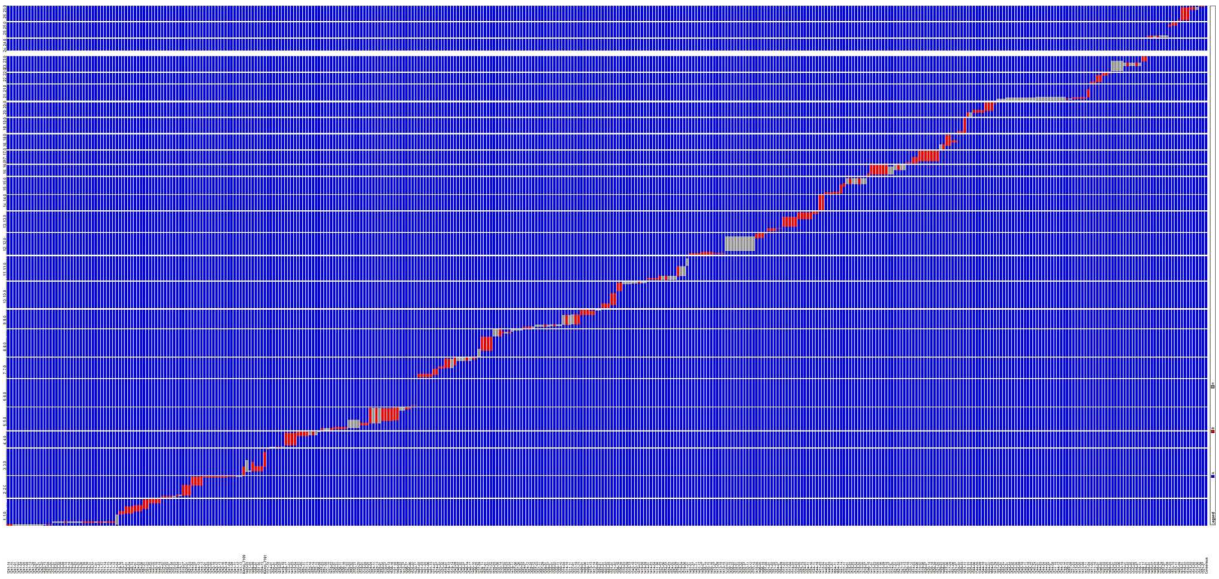

FIGURE 3 Graphical representation of the genomic composition of near-isogenic lines (NILs) in the Acala Maxxa background. Individual NILs are shown in the x-axis and the 26 chromosomes (sequentially from bottom to top) are represented in the y-axis. Blue blocks show homozygous alleles for the recurrent parent—Acala Maxxa. Red blocks show single introgressions of Pima S6 in homozygous state, while gray blocks show single introgressions of Pima S6 in heterozygous state. See Dataset S1 for details.

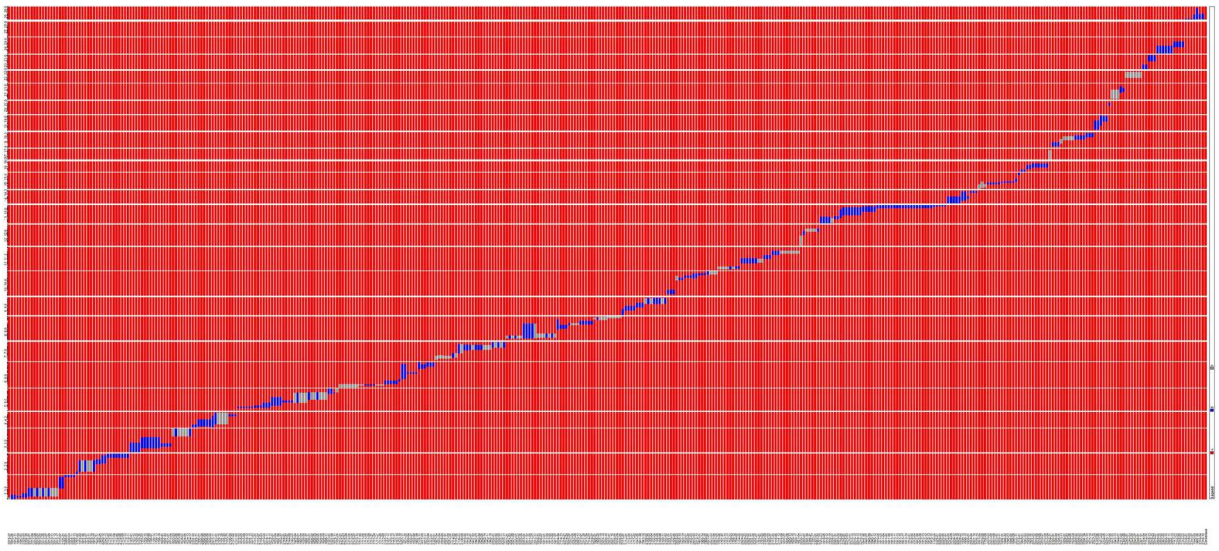

FIGURE 4 Graphical representation of the genomic composition of near-isogenic near-line (NILs) in the Pima S6 background. Individual NILs are shown in the x-axis, and the 26 chromosomes (sequentially from bottom to top) are represented in the y-axis. Red blocks show homozygous alleles for the recurrent parent—Pima S6. Blue blocks show single introgressions of Acala Maxxa in homozygous state, while gray blocks show single introgressions of Acala Maxxa in heterozygous state. See Dataset S2 for details.

TABLE 3 Performance comparison for models trained on successively larger datasets. The first row is the baseline model, trained only on UGA2020V imagery. The subsequent rows show the effects on mean average precision (mAP) of adding progressively more aerial images, annotated using the pseudo-active learning procedure described in section.

Dataset	No. of images	mAP@0.5	mAP@.5:.95
UGA2020V	1486	0.79	0.46
UGA2020V + aerial (1st round)	1646	0.78	0.45
UGA2020V + aerial (2nd round)	1902	0.86	0.53
UGA2020V + aerial (3rd round)	2142	0.86	0.54
UGA2020V + aerial (4th round)	2462	0.87	0.54

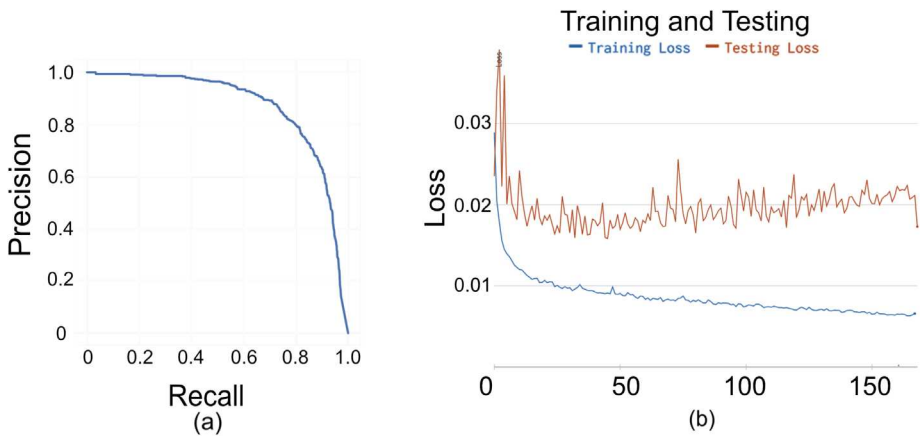

FIGURE 5 Model performance and training curves. The precision-recall curve (a) and loss curves (b) during a single training run of the model used for flower counting. Here, the model was trained for 160 epochs.

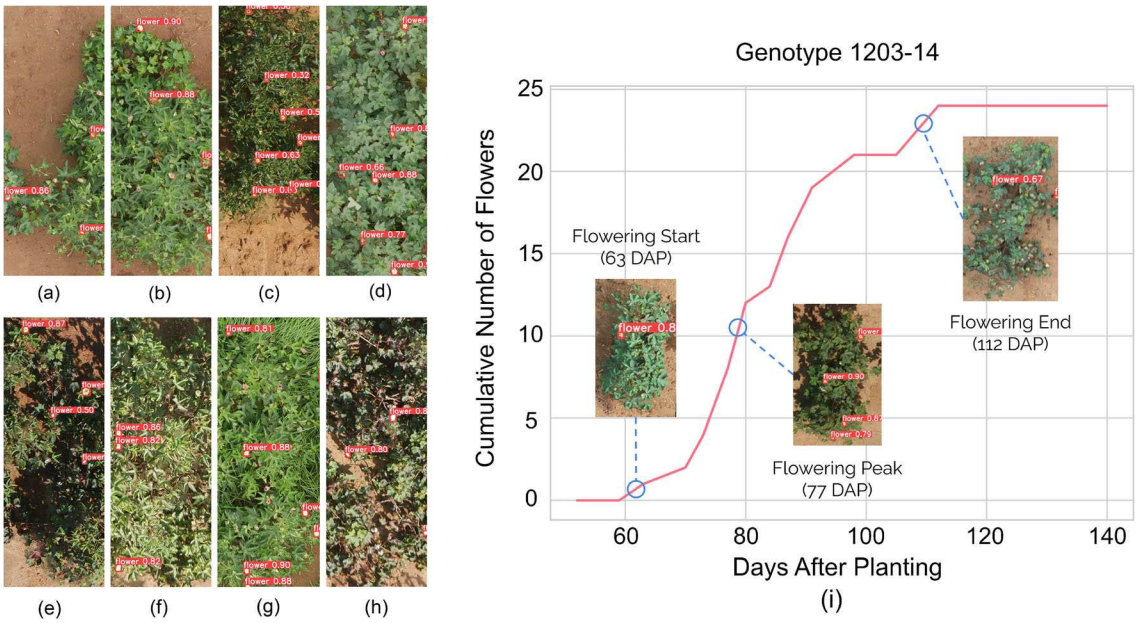

FIGURE 6 Qualitative examples of the bloom detection results. The model performs well even under heavy occlusion (a and b). It can handle Pima-like phenotypes (c) and only detects unpollinated flowers (d). It also performs reasonably even when the image quality is degraded due to poor lighting (e) or orthomosaicing artifacts (f). The presence of weeds does not degrade performance (g). Occasionally, the model confuses open bolls with flowers (h); however, this generally happens under conditions when it is difficult for even a human to distinguish them. Subfigure (i) shows a flowering curve for a single plot (5453) generated by the proposed analysis process. The corresponding plot images with annotated detections are shown at key points in the season.

found to be quite effective. This likely speaks to the high quality of the aerial imagery. Even so, the mAP@0.5 improved by nearly 12% from the addition of 976 annotated aerial images. It can be seen, however, that the effect rapidly saturates with additional data.

The deep learning model produces high-quality detections for aerial imagery (Figure 6). In general, we find that the model works about as well on all phenotypes and is able to distinguish unpollinated flowers from bolls and pollinated flowers. The most common failure case arises in late-season data, due to the presence of open bolls. The model will occasionally mistake such bolls for white flowers. This issue is exacerbated by poor lighting conditions that cause both bolls and flowers to be overexposed, eliminating many of their distinguishing features. It is worth pointing out that, under these conditions, even humans will struggle to differentiate between the two.

Overall, we are satisfied with the performance of the model, but additional steps could be taken to improve it even further. Most obviously, the training dataset could be expanded, including with data from additional growing seasons in order to enhance the model's generalization ability. Furthermore, comparing automatic counts with "ground truth" counts obtained by manually counting flowers in the field would enhance our confidence in the model. Unfortunately, those data are not available for this study, but we plan to collect them in follow-up work.

The remote sensing protocol, including the specific drone hardware and flight altitude, as well as environmental conditions such as

weather can have a significant impact on the flower analysis process. In order to mitigate this, we chose a strategy of employing a large and diverse training dataset for our counting model, validating on data from previous years in order to ensure generalizability. Even so, since this study focused on a single experimental field during a single season, we cannot guarantee that our approach is entirely free from bias. We plan to further validate our analysis pipeline by testing it on data from future seasons.

Figure 6i documents the flowering of a single plot over the entire season. The flowering curve exhibits a characteristic sigmoidal shape, with flowering starting at around 63 DAP and reaching its peak at 77 DAP. The flowering process diminishes around 112 DAP. Beyond this point, the plants have few flowers but many bolls.

3.3 | Correlation among flowering traits

The five flowering traits analyzed in this study show medium to no correlation with each other except for flowering duration, which has high correlations with flowering start time and end-time (Table 4). Although all five flowering-related traits were extracted from the same flower count data, the observation that these extracted traits are not highly correlated with each other not only suggests that our method does a decent job in defining the overall flowering pattern in cotton but also solidifies the results of our genetic analysis that co-localization of QTLs for different flowering-related traits is not merely due to high correlation among them.

TABLE 4 Correlation between flowering-related traits in reciprocal near-isogenic line (NIL) populations. The upper part of the matrix shows the correlation coefficients and the lower part shows the corresponding *p*-values.

Pima S6 population				
Peak	0.18	0.28	0.06	−0.01
0.0000	**Start**	−0.02	−0.75	−0.18
0.0000	0.5286	**End**	0.68	0.08
0.1247	0.0000	0.0000	**Duration**	0.19
0.7371	0.0000	0.0238	0.0000	**Slope**
Acala Maxxa population				
Peak	0.25	0.12	−0.07	−0.18
0.0000	**Start**	−0.05	−0.68	0.01
0.0012	0.1742	**End**	0.77	−0.02
0.0787	0.0000	0.0000	**Duration**	−0.02
0.0000	0.8606	0.6630	0.6654	**Slope**
Full data				
Peak	0.30	0.18	−0.09	−0.02
0.0000	**Start**	−0.03	−0.73	−0.04
0.0000	0.2597	**End**	0.70	0.04
0.0009	0.0000	0.0000	**Duration**	0.06
0.4566	0.1362	0.1265	0.0355	**Slope**

3.4 | Phenotypic performance of reciprocal NILs and parents

The cumulative flower counts over the 23 sessions of data collection for all the NILs are presented in Figure 7a. Individual lines and population mean for both backgrounds follow the expected sigmoidal pattern for cumulative flower counts. Curves for overall population mean as well as for individual lines within the respective population show that individual Acala Maxxa NILs produce flowers earlier than Pima S6 NILs, consistent with the flowering habit of the parents. While Acala Maxxa is relatively late flowering among Upland cottons and is considered a "full season" cultivar, *G. barbadense*-derived cottons such as Pima S6 are even later, despite being substantially lower yielding. An interesting observation seen in the study environment was that the NILs in the Pima S6 background produced more flowers over the growing period (Figure 7a), but the smaller size of the bolls in these lines ended up giving expected (lower than Acala Maxxa NILs) yields at harvest (data not shown). Additionally, many Pima S6 NILs were still flowering at the time of the first frost, potentially contributing to the higher number of flower counts in these lines.

One peculiar attribute of our individual NILs is the proportion of donor genome they carry. Since each NIL had only very small contributions (~0.9%) from the donor parent, the overall average of the respective populations reflected the phenotype of the recurrent parents at far higher replication than would otherwise have been practical, providing new information that informed our study. Thus, we utilized this unique, yet extremely significant, property of our populations to derive the phenotypes of the recurrent parents as the average phenotypic values of each background (averaged over 400 lines).

To further investigate the flowering response in these populations, we utilized the flower counts taken over the growing season and extracted phenotypes that quantified components of their flowering habits. These phenotypes include flowering start time, peak flowering time, flowering end-time, duration of flowering, and change in flowers per day (slope) during the inflection period defined as the duration from initiation of flowering to the peak flowering time of the lines and indicated by the exponential growth phase of the sigmoid curve. Figure 7b shows the distribution of flowering start time in the two reciprocal populations. In accord with the pattern seen in Figure 7a, Acala Maxxa NILs show significantly earlier flowering than Pima S6 NILs (Figure 8).

Figure 7c shows the distribution of flowering end-time in the two reciprocal populations. Flowering end-times were significantly different between the two populations (Figure 8), although the end-times show much less variation. Two factors might have played a role in this behavior. First, freezing temperatures on the second and third weeks of November might have stopped flowering. Second, owing to the reduced flowers produced by the plants following the first freeze, we stopped scanning the field after November 5, 2021. Nonetheless, the data we collected so far provided valuable information in observing the spread of flowering end-times in these populations in the 2–3 weeks span before we stopped collecting data. Figure 7d shows the distribution of peak flowering time in the reciprocal populations. As

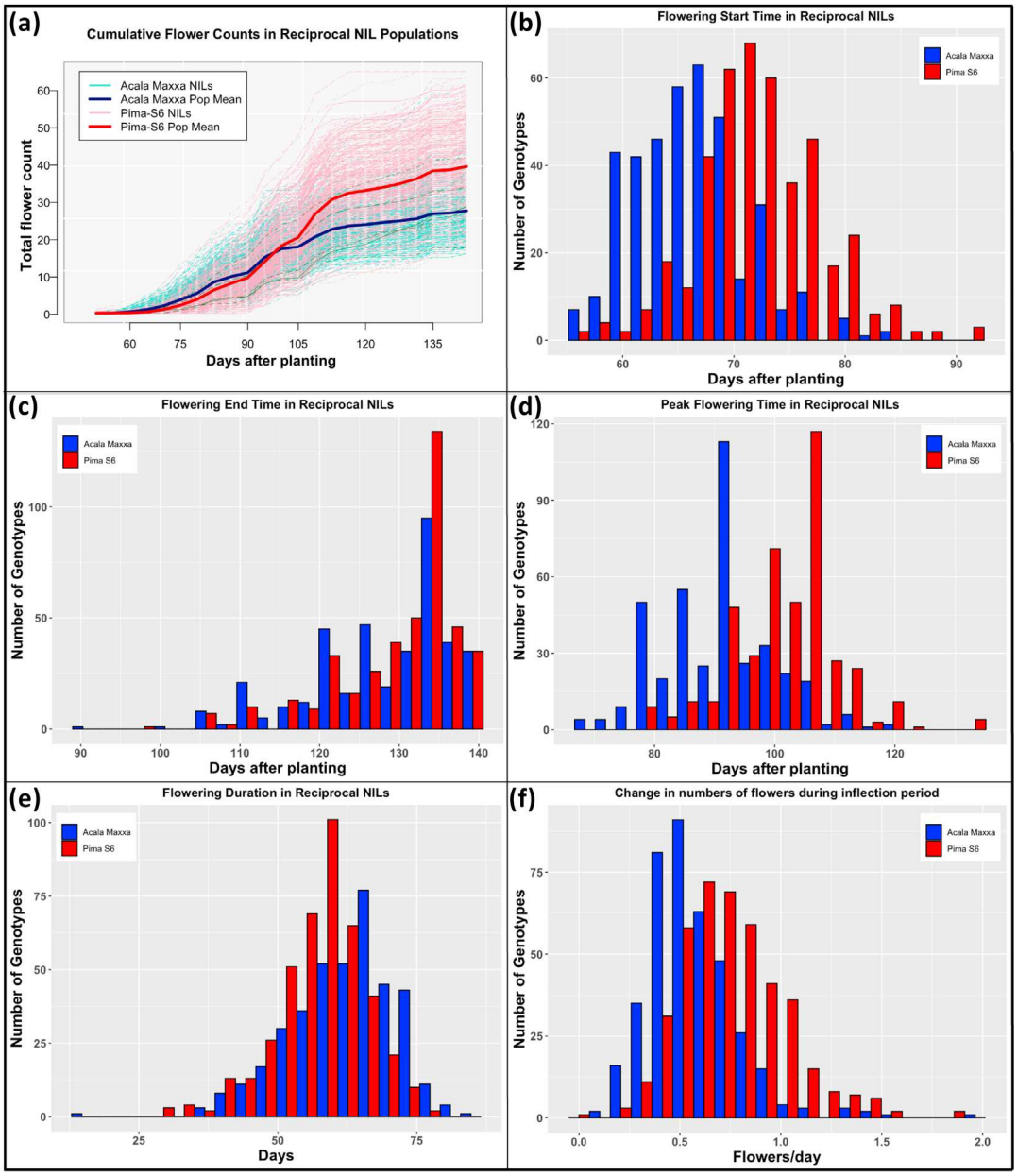

FIGURE 7 Distribution of flowering-related traits in the reciprocal set of near-isogenic lines (NILs). The traits include cumulative flower counts (a), flowering start time (b), flowering end-time (c), peak flowering time (d), flowering duration (e), and change in numbers of flowers during inflection period (f).

would be expected from the flowering trend graph and the flowering habit of the parents, Acala Maxxa NILs had significantly earlier peak flowering than Pima S6 NILs (Figure 8).

The average flowering duration of Acala Maxxa NILs was significantly shorter than that of Pima S6 NILs (Figure 7e), and the rate of increase in flowers per day was higher in Pima S6 NILs (Figure 7f). The population averages for both these traits were significantly different between the two reciprocal populations (Figure 8). The results presented in Figure 7b–f sum up the flowering pattern shown in Figure 7a. With shorter flowering duration and fewer flowers produced per day, the total flower counts in Acala Maxxa population lag Pima S6 population. On the other hand, longer growing season and a greater number of flowers produced per day boosts the total number of flowers in Pima S6 population albeit that they start flowering late in the season.

3.5 | Genetic control of floral trajectories

Marker-trait associations were carried out for all five traits related to flowering habit in the reciprocal set of NILs, and QTLs were identified following the procedure described in Section 2. A total of 24 QTLs were identified. QTLs for all five traits were identified in both backgrounds. The largest number of QTLs were identified for slope of

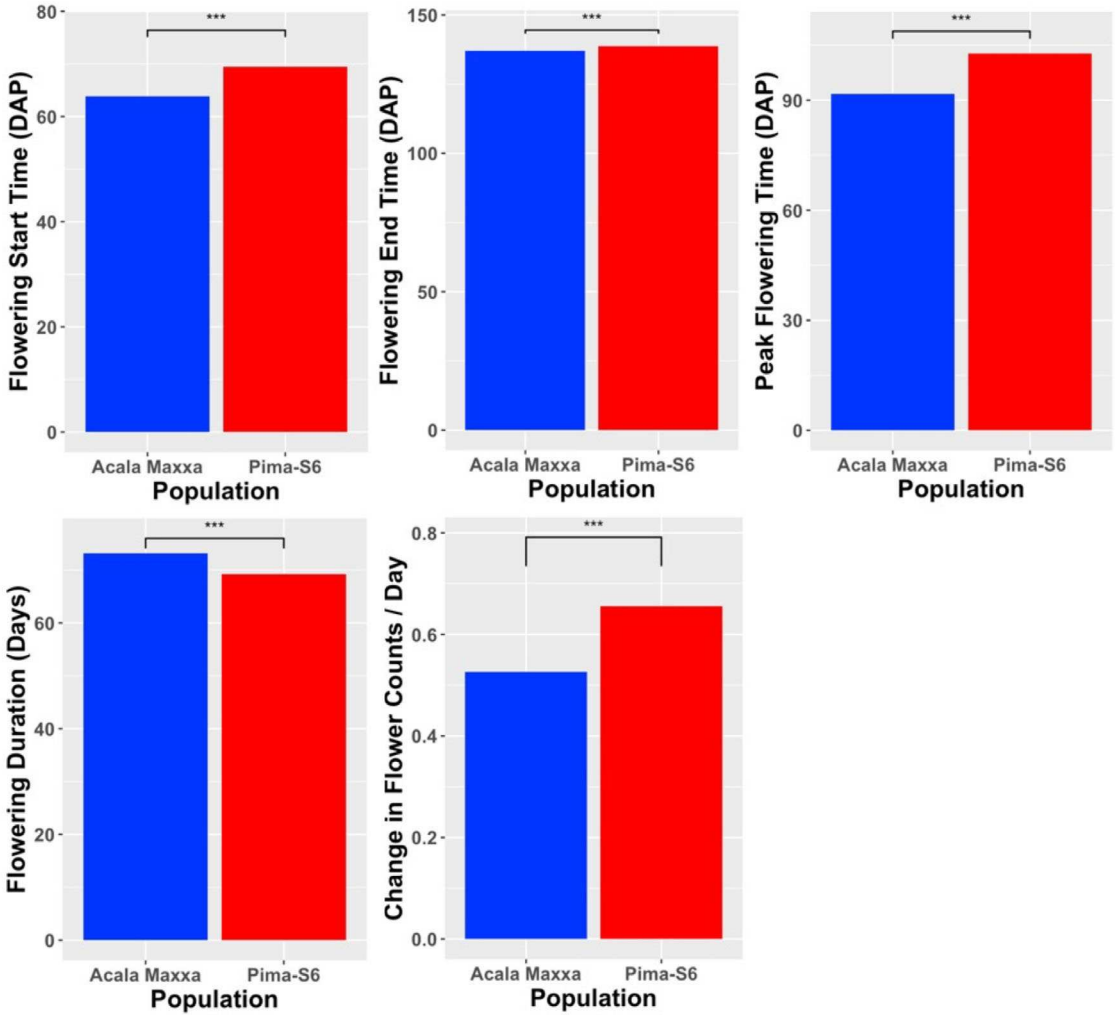

FIGURE 8 Population averages for flowering time related traits in the reciprocal set of NILs.(*, **, and *** significantly different at $\alpha = 0.05$, 0.01, and 0.001, respectively).

flowering progression in the Acala Maxxa background, while the smallest (one) were identified for peak flowering time in each of the two backgrounds. The phenotypic variance explained by these QTLs ranged from 2.21% to 15.94% (Table 4). Among the 24 QTLs identified for flowering-related traits in the reciprocal set of NILs, 19 were located in the At subgenome and five in the Dt subgenome.

Four genomic regions (QTLs) that influenced the season-long difference in growth trajectories of the two focal taxa (Table 5) exemplify factors that might be identified and manipulated in other perennials using iterative phenotyping made possible by aerial imagery and deep learning. In Pima S6 background, introgression of an Acala Maxxa chromosomal segment at a chromosome 3 region centered on the DNA marker S3_20315421 was associated with an advance of both the start and the end of the flowering period (consistent with the Acala Maxxa source of the introgressed segment) and a higher slope of the progression of flowering (resembling Pima S6 rather than Acala Maxxa), explaining 7%–16% of phenotypic variation in these traits. These traits showed simple additive genetics. In Acala Maxxa background, introgression of a Pima S6 chromosomal segment at a chromosome 2 region centered on the DNA marker S2_47565651 was associated with additive effects delaying flowering start time and decreasing the slope of its progression, explaining 4%–12% of phenotypic variation in these traits. Heterozygotes for the introgressed segment showed partial dominance (absolute dominance to additive effect ratio of 0.2 to 0.8) opposite of the additive effects for slope of flowering progression, outperforming the higher performing parent (Pima S6) for the number of flowers produced. Increased production of flowers and the slope of the progression of flowering in a high yielding background indicated that cotton hybrids for this genomic region could be earlier and higher yielding than the superior parent (Acala Maxxa)—perhaps contributing to the striking productivity motivating widespread cultivation of *G. hirsutum* × *Gossypium barbadense* F_1 hybrids in regions where low labor costs make hybrid seed production economical.

Another genomic region introgressed from Pima S6 into Acala Maxxa background on chromosome 8 (carrying DNA markers S8_15772489 and S8_2575405) was associated with additive effects of advancing the start of flowering and increased slope of floral

TABLE 5 Quantitative trait loci (QTLs) defining floral trajectory in a reciprocal set of near-isogenic lines. + indicates the additive effect of the QTL, # indicates the dominance effect of the QTL, and * indicates the gene action of the QTL defined as the ratio of dominance to additive effect. The logarithm of the odds (LOD) score and percent phenotypic variance explained (PVE) are provided for each QTL.

Trait	QTL name	Chr	Marker	LOD	PVE (%)	a	d	d/a
Pima S6 background								
Start time	qSDGb.03.01	3	S3_20315421	4.32	7.66	−4.37		
End-time	qEDGb.03.01	3	S3_20315421	4.88	15.94	−7.37		
	qEDGb.06.01	6	S6_90472087	3.07	6.32	6.23		
	qEDGb.19.01	19	S19_64167229	3.88	3.15	−38.52		
Peak time	qPDGb.07.01	7	S7_11615229	3.17	5.52	−15.05		
Duration	qDUGb.11.01	11	S11_27368710	3.06	4.91	−9.81		
	qDUGb.13.01	13	S13_30461025	4.51	3.22	−17.89		
Slope	qSLGb.03.01	3	S3_34861752	6.93	4.42	0.35		
	qSlGb.05.01	5	S5_23021069	3.28	2.96	0.08		
Acala Maxxa background								
Start time	qSDGh.08.01	8	S8_25754059	3.42	2.21	−4.39	24.85	−5.66
	qSDGh.02.01	2	S2_49415165	3.12	12.33	7.19		
	qSDGh.18.01	18	S18_34178947	3.13	4.22	19.51		
End-time	qEDGh.02.01	2	S2_6037287	4.38	4.97	−4.72	−32.38	6.86
	qEDGh.22.01	22	S22_30860588	3.65	4.93	9.74		
Peak time	qPDGh.03.01	3	S3_34521566	3.44	3.79	16.11		
Duration	qDUGh.02.01	2	S2_8696312	4.38	5.96	9.58	−29.41	−3.07
	qDUGh.13.01	13	S13_22314033	3.97	3.91	13.35		
Slope	qSLGh.01.01	1	S1_7462004	7.83	2.57	0.66	0.16	0.24
	qSLGh.02.01	2	S2_47565651	4.12	4.33	−0.02	0.33	−16.50
	qSLGh.08.01	8	S8_15772489	4.42	3.74	0.89		
	qSLGh.12.01	12	S12_8533077	6.55	3.27	0.14		
	qSLGh.13.01	13	S13_77978744	5.53	4.28	0.34		
	qSLGh.20.01	20	S20_10467201	4.11	9.63	0.94	0.38	0.40
	qSLGh.22.01	22	S22_37477542	5.191	6.13	0.39		

progression, explaining 2%–4% of the phenotypic variance. The heterozygotes carrying this introgression, however, delayed the flowering start time significantly, suggesting the use of fixed lines for this QTL to obtain early initiation of flowering and escape early frost during boll development stages, while producing more flowers in the high yielding Acala Maxxa background, towards ever higher yields.

One additional QTL shows more subtle effects on season-long growth trajectories. In Acala Maxxa background, introgression of a chromosomal segment from Pima S6 at a chromosome 12 region centered on the DNA marker S22_30860588 was associated with a delay of the end of the flowering period, and a nearby marker (S22_37477542) was associated with a higher slope of the progression of flowering, explaining 4%–6% of phenotypic variation in these traits. Additional QTLs on chromosomes 1, 2, 13, and 20 (in Acala Maxxa background) and 3 and 5 (Pima S6 background) had significant effects on the slope of the floral progression, reflecting whether a genotype is "conservative" (dispersing flowering over a long period, thus increasing the likelihood that at least a subset coincides with favorable conditions for fertilization) or "aggressive," flowering in a narrower window that may be optimal, resulting in more uniform and therefore more valuable seed and fibers (a seed byproduct, i.e., the primary economic product for which cotton is cultivated.)

3.6 | Reciprocal correspondence

A perplexing observation was the complete absence of reciprocally corresponding loci affecting the traits (QTLs). Under additive inheritance, a high degree of reciprocity is expected in the number and magnitude of effects of introgressed regions in reciprocal populations, suggesting that multi-locus epistasis may play a major role in the salient traits. Another factor may be the small phenotypic effects of most of the identified QTL, increasing the likelihood that one or both members of a reciprocal pair elude detection (Broman, 2001). However, one reciprocally introgressed segment in chromosome 3 was affecting peak flowering time in Pima S6 background, while it affected flowering start and end as well the slope of floral progression in the Acala Maxxa background. Given that these traits are correlated traits

extracted from the same set of flower count data, further scrutiny of this genomic regions might shade more light on the underlying mechanisms.

Close inspection of markers associated with traits at likelihoods below the LOD threshold of three identified five additional cases of loci within 3 Mb of their reciprocal location with opposite phenotypic effects in the reciprocal background. Noting that the NIL sets, respectively, cover 71.48% and 78.72% of the Acala Maxxa and Pima S6 genomes, about 1/3 of the genome lacks reciprocally introgressed segments—for example, a QTL for peak flowering time was identified in chromosome 24 in the Pima S6 background, which lacked introgression in the Acala Maxxa background.

3.7 | Similarity with QTLs previously reported

Flowering response in cotton is a very complex trait, and it has not been as comprehensively studied as fiber quality or yield traits. An additional level of complexity comes from the definition of flowering-related traits that are being extracted and utilized in genetic mapping. Some mapping studies have defined flowering-related traits as time of flowering initiation, flowering duration, and flowering cessation, while others have early maturity traits as synonymous with early flowering traits (Fu et al., 2019; Guo et al., 2009; Jia et al., 2016; Li et al., 2012; Li et al., 2017; Yufang et al., 2007). A few studies have also used time to 50% flowering and days to 50% boll opening as well as percentage of open flowers and open bolls to characterize early maturity in cotton (Adhikari, 2015; Jia et al., 2016).

Using an interspecific recombinant inbred line (RIL) population developed from the cross between an early maturing variety CCRI36 and a late maturing accession G2005, Jia et al. (2016) identified a total of 247 flowering-related QTLs, 10 of which (for days to 50% flowering) colocalized in the same chromosomes in which we identified QTLs for flowering start date and flowering duration (chromosomes 2, 3, 8, 11, and 18). In a study using 169 upland cotton backbone cultivars and breeding lines (Fu et al., 2019), GWAS identified two significant marker-trait associations for NFFB in chromosome 3, three in chromosome 8, and one in chromosome 18. The four QTLs for flowering start date identified in our study are each located in these same chromosomes.

Identification of QTL correspondence for flowering-related QTLs in cotton is a daunting task given the variation in trait names and methods used to collect and report these traits. Utilizing the hypergeometric probability method described by Feltus et al. (2006), we looked for correspondence to the flowering-related QTLs identified in our study with those previously reported. A total of 499 previously reported QTLs (17 publications from 1998 to 2021, Table S1) were used to identify correspondence to the QTLs identified in our study (Table 4). To account for the nonuniformity in the type of traits measured to quantify flowering behavior, traits of similar nature were used for correspondence. For example, QTLs for FT, NFFB/FFBN/NFB, and HFFBN, which were used to quantify the initiation of flowering in cotton, were used to correspond "Start Date" QTLs in our study. Likewise, bud period (BP), flower and bud period (FBP), and GP QTLs were used to find correspondence to flowering duration QTLs. Yield percent before frost (YPBF), QTLs were used as correspondence to flower end-time QTLs. No QTLs of corresponding nature were identified for peak flowering time and slope.

A total of five QTLs (related to three traits described in current study) corresponded nonrandomly ($p < 0.05$) with at least one of the previously reported QTLs (Table S2). For example, flowering start date QTL reported here showed nonrandom overlap ($p = 0.0003$) with NFFB QTL reported by Fu et al. (2019). Similarly, flowering duration QTL identified in our study corresponded with FBP QTL reported by Su et al. (2016) in a natural population of cotton. Likewise, three nonrandomly overlapping QTLs were identified for flowering end date (Fu et al., 2019; Li et al., 2012, 2016). These few, but statistically significant, nonrandomly correspondent QTLs provide for validation of the QTLs identified in our study as well as for the need of further exploration of these genomic regions to dissect these important agroeconomic traits in cotton.

The four major effect flowering-related QTLs identified in this study are on chromosomes 3, 11, 12, and 13. Numerous published studies show that these chromosomes also host QTLs for fiber quality traits (Rong et al., 2007; Said et al., 2013; Wang et al., 2013). Guo et al. (2008) also identified QTLs for NFB clustered with fiber quality QTLs and suggested that most NFB QTLs were linked with some QTLs related to fiber quality. In addition, several other fiber quality QTL hotspots previously identified might harbor the small-effect QTLs identified in our study. If there is a positive linkage between flowering-related and fiber quality QTLs co-occurring in these clusters, selection of genotypes with favorable flowering behavior would also select fiber and yield genes/gene combinations from the donor parent. Simultaneous selection of genotypes that flower early and have short flowering duration with desired fiber quality traits by marker-assisted selection may accelerate cotton improvement. This demands further exploration of these immortal lines for relationship between flowering-related traits and fiber quality-related traits. Investigation into these relationships could be one of our next adventures in further deciphering the genomic organization of these QTL hotspots in the cotton genome.

3.8 | Identification of candidate gene(s)

Mapping traits to small regions and deducing candidate genes potentially affecting these traits are of especially high value in woody perennials with complex genomes and long generation times that slow breeding progress. For example, on chromosome 2 near S2_6037287, there was *ERF1B* (*Gh_A02G0377*), which is expressed during the four-leaf visible stage and flowering stage and is responsible for regulation of transcription- and ethylene-activated signaling (mostly defense related) during flowering, fruit development, and ripening in woody plants (El-Sharkawy et al., 2009; Ninh et al., 2021; Zhang, Zhang, et al., 2021). Another gene, *PRP* (*Gh_A02G0378* and *Gh_A02G0379*), in close proximity to *ERF1B*, has similar functions and expression

patterns and has important roles in restructuring cell walls during cotton fiber elongation (Feng et al., 2004). Similarly, *GATA15* (*Gh_A121163*) was identified as the nearest gene of interest on chromosome 12. This gene is responsible for developmental growth and response to temperature stimulus and is shown to be expressed during flowering and petal differentiation stages in flower, flower pedicel, and inflorescence meristem (Zhang et al., 2019).

While all these genes are potential candidates for flowering response genes in cotton, further investigation is necessary to verify their role in floral initiation and flowering habit in cotton. Genome-wide association approaches taking advantage of whole-genome or reduced representation sequencing may afford even finer map precision than the NIL approach in natural populations with long histories of recombination, albeit requiring careful measures to mitigate false-positive associations.

4 | CONCLUSION

Iteratively tracking flowering over a long season by applying an efficient flower detection and counting deep learning model to a well-structured mapping population has identified both additive and transgressive genetic factors conferring different reproductive strategies that might be successful in different environments. The reciprocal set of NILs we developed not only provides a well-suited set of individuals for efficient mapping of these traits but also helps investigate the reciprocal effects of the introgressed segments into the recipient genome. All the flowering traits used in this study showed significant differences between genotypes, indicating that, albeit imperfect, the counting approach is able to measure genetically controlled differences in flowering attributes. The counting-by-detection approach for identifying and collecting flower counts in cotton could likely be applied to other crops with similar canopy structures, provided that individual flowers can be distinguished by aerial imaging. While the performance of the counting model can be improved, the partially automated annotation approach makes it routine to expand the dataset as required to achieve acceptable performance on other taxa.

AUTHOR CONTRIBUTIONS

Jeevan Adhikari: Methodology; investigation; data curation; writing—original draft; writing—review and editing; visualization. **Daniel Petti:** Methodology; software; investigation; data curation; writing—original draft; writing—review and editing; visualization. **Deepak Vitrakoti:** Investigation. **Wiriyanat Ployaram :** Investigation. **Changying Li:** Conceptualization; resources; writing—review and editing; supervision; project administration; funding acquisition. **Andrew H. Paterson**: Conceptualization; resources; writing—review and editing; supervision; project administration; funding acquisition.

ACKNOWLEDGEMENTS

The authors acknowledge Rui Xu for helping construct the data collection platform and collecting some initial data, as well as Gary Pierce for his help managing the experimental field.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The genomic fingerprinting data of the BC_5F_1 lines has been deposited in the Dryad repository (https://doi.org/10.5061/dryad.kh189329v). The genomic composition of the two reciprocal NIL populations and the plot wise metrices (flowering-related data obtained from the flower counting model) used in the analysis are included as Supporting Information. The annotated data used to train the model is available at https://doi.org/10.5061/dryad.5qfttdzhb.

ORCID

Jeevan Adhikari https://orcid.org/0000-0002-3815-3067
Daniel Petti https://orcid.org/0000-0002-5293-445X
Deepak Vitrakoti https://orcid.org/0000-0001-9009-364X
Wiriyanat Ployaram https://orcid.org/0000-0002-8429-4098
Changying Li https://orcid.org/0000-0003-2590-4797

REFERENCES

Adak, A., Murray, S. C., Anderson, S. L., Popescu, S. C., Malambo, L., Romay, M. C., & de Leon, N. (2021). Unoccupied aerial systems discovered overlooked loci capturing the variation of entire growing period in maize. *The Plant Genome*, *14*, e20102. https://doi.org/10.1002/tpg2.20102

Adamsen, F. a. (2000). Method for using images from a color digital camera to estimate flower number. *Crop Science*, *40*, 704–709. https://doi.org/10.2135/cropsci2000.403704x

Adhikari, J. (2015). Effects of exotic genotypes and genetic backgrounds on fiber quality and plant architectural traits in upland cotton (*Gossypium hirsutum* L.).

Adhikari, J., Chandnani, R., Vitrakoti, D., Khanal, S., Ployaram, W., & Paterson, A. H. (2023). Comparative transmission genetics of introgressed chromatin in reciprocal advanced backcross populations in Gossypium (cotton) polyploids. *Heredity*, *130*, 209–222. https://doi.org/10.1038/s41437-023-00594-w

Akiva P. (2020). Finding berries: segmentation and counting of cranberries using point supervision and shape priors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Bălan, L., Kiyo, K. K., Deriabin, D., Hoang, L., Ivaniuk, A., & Dada, Y. (2020). dr3s. quantumblacklabs/kedro: 0.17.0. quantumblacklabs/kedro: 0.17.0. Zenodo.

Bellocchio, E. A. (2019). Weakly supervised fruit counting for yield estimation using spatial consistency. *IEEE Robotics and Automation Letters*, *4*, 2348–2355. https://doi.org/10.1109/LRA.2019.2903260

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl*. Springer.

Broman, K. W. (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal*, *30*(7), 44–52.

Cohen, J. P. (2017). Count-ception: counting by fully convolutional redundant counting. Proceedings of the IEEE international conference on computer vision (ICCV) workshops.

El-Sharkawy, I., Sherif, S., Mila, I., Bouzayen, M., & Jayasankar, S. (2009). Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. *Journal of Experimental Botany*, *60*, 907–922. https://doi.org/10.1093/jxb/ern354

Feltus, F. A., Hart, G. E., Schertz, K. F., Casa, A. M., Kresovich, S., & Abraham, S. Paterson, A. H.(2006). Alignment of genetic maps and

QTLs between inter- and intra-specific sorghum populations. *Theoretical and Applied Genetics*, *112*, 1295–1305. https://doi.org/10.1007/s00122-006-0232-3

Feng, J.-X., Ji, S.-J., Shi, Y.-H., Xu, Y., Wei, G., & Zhu, Y.-X. (2004). Analysis of five differentially expressed gene families in fast elongating cotton fiber. *Acta Biochimica et Biophysica Sinica*, *36*, 51–56. https://doi.org/10.1093/abbs/36.1.51

Fankhauser, C., & Chory, J. (1997). Light control of plant development. *Annual Review of Cell and Developmental Biology*, *13*(1), 203–229. https://doi.org/10.1146/annurev.cellbio.13.1.203

Fu, Y., Dong, C., Wang, J., Wang, Y., & Li, C. (2019). Genome-wide association study reveals the genetic control underlying node of the first fruiting branch and its height in upland cotton (*Gossypium hirsutum* L.). *Euphytica*, *215*, 35. https://doi.org/10.1007/s10681-019-2361-1

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., & Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics*, *2019*, 1525874. https://doi.org/10.34133/2019/1525874

Giuffrida, M. V. (2019). Leaf counting without annotations using adversarial unsupervised domain adaptation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2590-2599.

Grover, C. E., Gallagher, J. P., & Wendel, J. F. (2015). Candidate gene identification of flowering time genes in cotton. *The Plant Genome*, *8*, plantgenome2014.12.0098. https://doi.org/10.3835/plantgenome2014.12.0098

Guo, Y., McCarty, J. C., Jenkins, J. N., An, C., & Saha, S. (2009). Genetic detection of node of first fruiting branch in crosses of a cultivar with two exotic accessions of upland cotton. *Euphytica*, *166*, 317–329. https://doi.org/10.1007/s10681-008-9809-z

Guo, Y., McCarty, J. C., Jenkins, J. N., & Saha, S. (2008). QTLs for node of first fruiting branch in a cross of an upland cotton, *Gossypium hirsutum* L., cultivar with primitive accession Texas 701. *Euphytica*, *163*, 113–122. https://doi.org/10.1007/s10681-007-9613-1

Herr, A. W. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, *63*, 1722–1749. https://doi.org/10.1002/csc2.21028

Jia, X., Pang, C., Wei, H., Wang, H., Ma, Q., Yang, J., & Yu, S. (2016). High-density linkage map construction and qtl analysis for earliness-related traits in *Gossypium hirsutum* L. *BMC Genomics*, *17*, 909. https://doi.org/10.1186/s12864-016-3269-y

Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, *2020*, 4152816. https://doi.org/10.34133/2020/4152816

Jiang, Y., Li, C., Xu, R., Sun, S., Robertson, J. S., & Paterson, A. H. (2020). DeepFlower: A deep learning-based approach to characterize flowering patterns of cotton plants in the field. *Plant Methods*, *16*, 156. https://doi.org/10.1186/s13007-020-00698-y

Kohel, R. J., Richmond, T. R., & Lewis, C. F. (1974). Genetics of flowering response in cotton. VI. Flowering behavior of *Gossypium hirsutum* L. and *G. barbadense* L. hybrids1. *Crop Science*, *14*, 696–699.

Kushanov, F. N., Turaev, O. S., Ernazarova, D. K., Gapparov, B. M., Oripova, B. B., Kudratova, M. K., Rafieva, F. U., Khalikov, K. K., Erjigitov, D. S., Khidirov, M. T., & Kholova, M. D. (2021). Genetic diversity, qtl mapping and MAS technology in cotton (Gossypium spp.). *Frontiers in Plant Science*, *12*, 2971.

Leal Valentim, F., Mourik, S.v, Posé, D., Kim, M. C., Schmid, M., van Ham, R. C., Busscher, M., Sanchez-Perez, G. F., Molenaar, J., Angenent, G. C., Immink, R. G., & van Dijk, A. D. (2015). A quantitative and dynamic model of the Arabidopsis flowering time gene regulatory network. *PloS one*, *10*(2), e0116973. https://doi.org/10.1371/journal.pone.0116973

Lewis, C. F., & Richmond, T. R. (1960). The genetics of flowering response in cotton. II. Inheritance of flowering response in a *Gossypium barbadense* cross. *Genetics*, *45*, 79–85. https://doi.org/10.1093/genetics/45.1.79

Li, B., Chen, L., Sun, W., Wu, D., Wang, M., Yu, Y., & Yang, X. (2020). Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton. *Plant Biotechnology Journal*, *18*, 2533–2544. https://doi.org/10.1111/pbi.13431

Li, B., Huang, H., Zhang, A., Liu, P., & Liu, C. (2021). Approaches on crowd counting and density estimation: A review. *Pattern Analysis and Applications*, *24*, 853–874. https://doi.org/10.1007/s10044-021-00959-z

Li, C., Wang, C., Dong, N., Wang, X., Zhao, H., Converse, R., & Wang, Q. (2012). Qtl detection for node of first fruiting branch and its height in upland cotton (*Gossypium hirsutum* L.). *Euphytica*, *188*, 441–451. https://doi.org/10.1007/s10681-012-0720-2

Li, C., Wang, X., Dong, N., Zhao, H., Xia, Z., Wang, R., & Wang, Q. (2013). Qtl analysis for early-maturing traits in cotton using two upland cotton (*Gossypium hirsutum* L.) crosses. *Breeding Science*, *63*, 154–163. https://doi.org/10.1270/jsbbs.63.154

Li, C., Zhang, J., Hu, G., Fu, Y., & Wang, Q. (2016). Association mapping and favorable allele mining for node of first fruiting/sympodial branch and its height in upland cotton (*Gossypium hirsutum* L.). *Euphytica*, *210*, 57–68. https://doi.org/10.1007/s10681-016-1697-z

Li, L., Zhao, S., Su, J., Fan, S., Pang, C., Wei, H., & Yu, S. (2017). High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). *PLoS ONE*, *12*, e0182918. https://doi.org/10.1371/journal.pone.0182918

Loladze, A., Rodrigues, F. A., Petroli, C. D., Muñoz-Zavala, C., Naranjo, S., Vicente, F. S., & Martini, J. W. (2024). Use of remote sensing for linkage mapping and genomic prediction for common rust resistance in maize. *Field Crops Research*, *308*, 109281. https://doi.org/10.1016/j.fcr.2024.109281

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: Counting maize tassels in the wild via local counts regression network. *Plant Methods*, *13*, 79. https://doi.org/10.1186/s13007-017-0224-0

Madec, S. a. (2019). Ear density estimation from high resolution RGB imagery using deep learning techniques. *Agricultural and Forest Meteorology*, *264*, 225–234 264. https://doi.org/10.1016/j.agrformet.2018.10.013

Marsden, M., McGuinness, K., Little, S., Keogh, C. E., & O'Connor, N. E. (2018). People, penguins and petri dishes: adapting object counting models to new visual domains and object types without forgetting. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE.

Ninh, T. T., Gao, W., Trusov, Y., Zhao, J.-R., Long, L., Song, C.-P., & Botella, J. R. (2021). Tomato and cotton G protein beta subunit mutants display constitutive autoimmune responses. *Plant Direct*, *5*, e359. https://doi.org/10.1002/pld3.359

Ogawa, D., Sakamoto, T., Tsunematsu, H., Kanno, N., Nonoue, Y., & Yonemaru, J.-i. (2020). Haplotype analysis from unmanned aerial vehicle imagery of rice MAGIC population for the trait dissection of biomass and plant architecture. *Journal of Experimental Botany*, *72*, 2371–2382. https://doi.org/10.1093/jxb/eraa605

Petti, D., & Li, C. (2022). Weakly-supervised learning to automatically count cotton flowers from aerial imagery. *Computers and Electronics in Agriculture*, *194*, 106734. https://doi.org/10.1016/j.compag.2022.106734

QGIS Development Team. (2009). Qgis geographic information system.

Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE.

Rong, J., Feltus, F. A., Waghmare, V. N., Pierce, G. J., Chee, P. W., Draye, X., & Paterson, A. H. (2007). Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development.

Genetics, *176*, 2577–2588. https://doi.org/10.1534/genetics.107.074518

Said, J. I., Lin, Z., Zhang, X., Song, M., & Zhang, J. (2013). A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics*, *14*, 776. https://doi.org/10.1186/1471-2164-14-776

Sekachev, B., Zhavoronkov, A., & Manovich, N. (2019). Computer Vision Annotation Tool: A Universal Approach to Data Annotation. Tech. rep.

Spindel, J. E., Dahlberg, J., Colgan, M., Hollingsworth, J., Sievert, J., Staggenborg, S. H., & Vogel, J. P. (2018). Association mapping by aerial drone reveals 213 genetic associations for *Sorghum bicolor* biomass traits under drought. *BMC Genomics*, *19*, 679. https://doi.org/10.1186/s12864-018-5055-5

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., & Yu, S. (2016). Identification of favorable snp alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*, *17*, 687. https://doi.org/10.1186/s12864-016-2875-z

Tan, C., Li, C., He, D., & Song, H. (2022, January). Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. *Computers and Electronics in Agriculture*, *193*, 106683. https://doi.org/10.1016/j.compag.2021.106683

Tan, C., Li, C., He, D., & Song, H. (2023). Anchor-free deep convolutional neural network for tracking and counting cotton seedlings and flowers. *Computers and Electronics in Agriculture*, *215*, 108359. https://doi.org/10.1016/j.compag.2023.108359

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, *5*, 1680–1716. https://doi.org/10.3390/make5040083

Thorp, K. R. (2016). Lesquerella seed yield estimation using color image segmentation to track flowering dynamics in response to variable water and nitrogen management. *Industrial Crops and Products*, *86*, 186–195. https://doi.org/10.1016/j.indcrop.2016.03.035

Waddle, B. M., Lewis, C. F., & Richmond, T. R. (1961). The genetics of flowering response in cotton. Iii. Fruiting behaviour of **Gossypium hirsutum** race lati-folium in a cross with a variety of cultivated american upland cotton. *Genetics*, *46*, 427–437. https://doi.org/10.1093/genetics/46.4.427

Wang, C. C., Chang, P.-C., Ng, K.-L., Chang, C.-M., Sheu, P. C., & Tsai, J. J. (2014). A model comparison study of the flowering time regulatory network in Arabidopsis. *BMC Systems Biology*, *8*(1), 15. https://doi.org/10.1186/1752-0509-8-15

Wang, X., Zhang, R., Song, W., Han, L., Liu, X., Sun, X., Luo, M., Chen, K., Zhang, Y., Yang, H., Yang, G., & Zhao, J. (2019). Dynamic plant height qtl revealed in maize through remote sensing phenotyping using a high-throughput unmanned aerial vehicle (UAV). *Scientific Reports*, *9*, 3458. https://doi.org/10.1038/s41598-019-39448-z

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A whole-genome DNA marker map for cotton based on the D-genome sequence of Gossypium raimondii L. *G3: Genes, Genomes, Genetics*, *3*, 1759–1767. https://doi.org/10.1534/g3.113.006890

Wei, G., Zheng, B., Potgieter, A., Diot, J., Watanabe, K., Noshita, K., Jordan, D. R., Wang, X., Watson, J., Ninomiya, S., & Chapman, S. (2018). Aerial imagery analysis—Quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Frontiers in Plant Science*, *9*, 1544.

Wu, X., Feng, H., Wu, D., Yan, S., Zhang, P., Wang, W., & Terzaghi, W. (2021). Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biology*, *22*, 185. https://doi.org/10.1186/s13059-021-02377-0

Xiong, H. a.-G. (2019). TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods*, *15*, 150. https://doi.org/10.1186/s13007-019-0537-2

Xu, R., Li, C., Paterson, A. H., Jiang, Y., Sun, S., & Robertson, J. S. (2018). Aerial images and convolutional neural network for cotton bloom detection. *Frontiers in Plant Science*, *8*, 2235. https://doi.org/10.3389/fpls.2017.02235

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., & Xiong, L. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications*, *5*, 5087. https://doi.org/10.1038/ncomms6087

Yufang, G., McCarty, J. C., Saha, S., Jenkins, J. N., & Jixiang, W. (2007). Qtl analyses of flowering time in a primitive cotton accession (**Gossypium hirsutum** L.). (pp. 231–231).

Zhang, J., Jia, X., Guo, X., Wei, H., Zhang, M., Wu, A., & Wang, H. (2021). Qtl and candidate gene identification of the node of the first fruiting branch (NFFB) by QTL-seq in upland cotton (**Gossypium hirsutum** L.). *BMC Genomics*, *22*, 882. https://doi.org/10.1186/s12864-021-08164-2

Zhang, J., Zhang, P., Huo, X., Gao, Y., Chen, Y., Song, Z., & Zhang, J. (2021). Comparative phenotypic and transcriptomic analysis reveals key responses of upland cotton to salinity stress during postgermination. *Frontiers in Plant Science*, *12*, 639104. https://doi.org/10.3389/fpls.2021.639104

Zhang, Z., Zou, X., Huang, Z., Fan, S., Qun, G., Liu, A., & Yuan, Y. (2019). Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of *Gossypium* genus. *Gene*, *680*, 72–83. https://doi.org/10.1016/j.gene.2018.09.039

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Adhikari, J., Petti, D., Vitrakoti, D., Ployaram, W., Li, C., & Paterson, A. H. (2025). Characterizing season-long floral trajectories in cotton with low-altitude remote sensing and deep learning. *Plants, People, Planet*, *7*(6), 1657–1673. https://doi.org/10.1002/ppp3.10644