

<https://doi.org/10.1038/s44264-025-00074-0>

Genetic mapping for agronomic, nutritional, and leaf vein traits in the indigenous crop *Gynandropsis gynandra*



Conor J. C. Simpson^{1,3}, Dêdi E. O. Sogbohossou², Gregory Reeves^{1,4}, M. Eric Schranz², Pallavi Singh^{1,5} ✉ & Julian M. Hibberd¹ ✉

Gynandropsis gynandra is a nutrient-rich, climate-resilient, C₄ under-utilised crop with potential to enhance food security in low-input farming systems. To support its genetic improvement, we performed linkage mapping using two F₂ populations (213 and 187 individuals) derived from Malaysian and Malawian founder lines. Under controlled greenhouse conditions, populations were phenotyped for agronomic traits (plant height, leaf area, flowering time), nutritional content (carotenoids, tocopherols), and anatomical features linked to C₄ photosynthesis (vein density, bundle sheath size). High-density SNP genotyping enabled construction of linkage maps and identification of 15 QTL. Shared QTL for plant size and flowering time across both populations suggest stable genetic control suitable for marker-assisted selection. Additional QTL for vein density and vitamin content provide insights into the genetic basis of agronomic traits and C₄ physiology. These results offer foundational tools for *G. gynandra* pre-breeding and reinforce its potential as a model for C₄ photosynthesis research and sustainable agriculture.

At least 250,000 plant species have been identified as edible. Of these, 7000 have been consumed by humans but in 1995 only twelve crops and five animals made up 75% of the world's food. Moreover, at this time 60% of global calories were provided by just three cereals, rice (*Oryza sativa*), maize (*Zea mays*) and wheat (*Triticum aestivum*)¹. This bottleneck in agricultural diversity led to interest in underutilised, neglected, or so-called orphan crops² however, there is debate about the use of such terms³ and now more inclusive and positive terms such as “opportunity” and “indigenous” crops are increasingly being used^{4,5}. Greater use of such species could reduce the demand to adjust conditions to suit the current limited portfolio of major crops and instead allow crops better adapted to local conditions to be deployed⁶. Thus, increased use of locally adapted species could increase resilience of supply chains and so contribute to improving the lives of the ~10% of the global population who are undernourished⁷. Better adoption of these species could also help agriculture cope with increasingly unpredictable extreme weather and greater climatic uncertainty that are applying mounting pressure on the global food system^{8,9}. Neglected and opportunity crops are defined as those that have been locally cultivated from sites at which the crop originates, or have become naturalised but been neglected by the global agricultural community^{10,11}. Orphan crops are often defined as

being rich in nutrient content, are climate resilient, locally available, and economically viable^{12–14}. Orphan crops have already been adopted, with examples including quinoa (*Chenopodium quinoa*)¹⁵ and India's targeting breeding programme of pearl millet (*Pennisetum glaucum*)¹⁶.

Gynandropsis gynandra (2n = 34) of the Cleomaceae family is a relative of *Arabidopsis* and Brassica crops in its sister family Brassicaceae. It has many common names including Shona Cabbage, Spider Plant, and Cat's Whiskers. It is grown as an indigenous vegetable crop in parts of Sub-Saharan Africa and Asia and matches the four criteria outlined above¹⁷. It has recently been recognised as being of high relevance to achieve food security in Africa due to its nutritional content and modelled ability to withstand climate change in Sub-Saharan Africa¹⁸. *Gynandropsis gynandra* has been identified as a strong candidate for improvement due to its high vitamin C (ascorbic acid) and β-carotene content along with high levels of phosphorus, potassium, calcium, iron, zinc, phenols and flavonoids compared with commercial cultivars of *Brassica oleracea* (var. *capitata* cv. Drumhead) and *Beta vulgaris* (L. cv. Fordhook Giant)¹⁹. Its seeds are high in polyunsaturated fatty acids and are cultivated in Asia for oil²⁰. The active ingredient of PIXALIA (Laboratoires Expanscience) is extracted from *G. gynandra* leaves for acne treatment, and *G. gynandra* has been identified as

¹University of Cambridge, Department of Plant Sciences, Cambridge, United Kingdom. ²Wageningen University & Research, Wageningen, The Netherlands.

³Present address: Karlsruhe Institute für Technologie Campus-Alpin IMKIFU, Garmisch-Partenkirchen, Germany. ⁴Present address: The Kiwifruit Breeding Centre Limited, Motueka, New Zealand. ⁵Present address: School of Life Sciences, University of Essex, Colchester, United Kingdom.

✉ e-mail: pallavi.singh@essex.ac.uk; jmh65@cam.ac.uk

being of medicinal interest with leaves high in flavonoids such as quercetin and kaempferol²¹ that are thought to have antithrombogenic properties²². *Gynandropsis gynandra* extracts have also been reported to contain moderate antifungal and antibacterial activities²³ and high glucosinolate content, which provides resistance to pests²¹. Lastly, its seeds are well suited for animal feed²⁰.

In addition to its practical applications, *G. gynandra* is a model organism for the study of C_4 photosynthesis. As such it has been used extensively as a system through which genetic determinants of C_4 gene expression have been investigated^{24–28}. Furthermore, a recently published genome sequence for *G. gynandra* confirmed its genome size of 740 Mb, and highlighted its utility in studying genome evolution²⁹.

Gynandropsis gynandra is grown across a wide geographic range that includes all inhabited continents²⁰, and these divergent environments and populations have resulted in a diverse gene pool. For example, resequencing data for fifty-three accessions identified three genetically distinct groups derived from West Africa, Asia and East/Southern Africa, respectively³⁰. These accessions exhibited natural variation in multiple economically important traits, including plant height, flowering time, and vitamin content³¹. Local cultivation is a major part of *G. gynandra*'s natural history, as it has been traditionally used by farmers who have independently selected for desirable traits such as stem and leaf colour, plant height, and flowering time³². Geographically diverse lines of *G. gynandra* can hybridize, but it is also self-compatible¹⁷. There is now a rich germplasm collection, with 471 accessions stocked by the World Vegetable Center (<https://avrdc.org/>). These factors make *G. gynandra* an excellent candidate for pre-breeding methods, such as identifying Quantitative Trait Loci (QTL) for use in Marker Assisted Selection (MAS).

Crossing highly divergent lines enables segregating mapping populations from which QTL can be identified, providing insight into trait inheritance and targets for marker assisted selection to enhance germplasm³³. To initiate this approach in *G. gynandra* we generated two F_2 mapping populations in 2018 and 2019 from parental lines of Malaysian and Malawian origin. These were assessed for agronomic traits (e.g. plant height, leaf area, flowering time), nutritional compounds (carotenoids and tocopherols), and leaf anatomical features relevant to C_4 photosynthesis.

Despite the growing recognition of *G. gynandra*'s value, no published studies to date have reported genetic linkage or QTL mapping in this species. This study therefore represents the first effort to dissect the genetic basis of trait variation in *G. gynandra* through bi-parental QTL mapping. We hypothesised that natural genetic variation between the founder lines would enable identification of heritable loci controlling key traits, including those underpinning the C_4 pathway.

Although photosynthesis has not been a major focus for crop improvement³⁴, it has been noted that selecting solely for yield can overlook potential avenues for crop improvement such as photosynthetic enhancement³⁵. Since the founders of Wag19 showed differences in traits underpinning the efficient C_4 photosynthesis pathway³⁶, we assessed this population for variation in vein density, which is crucial for this pathway³⁷. We anticipate that this work provides foundational tools for future marker-assisted breeding and will contribute to the development of improved *G. gynandra* germplasm.

Results

Phenotypic and heritability assessment for Wag18

Founders of the Wag18 population demonstrated statistically significant differences for all phenotypes assessed ($p < 0.05$, Supplementary Table S1). With the exception of alpha-tocopherol, all other traits had higher values in Malawi-02 than Malaysia-03 (Fig. 1). All traits displayed transgressive segregation compared with the founders (Fig. 1) and broad-sense heritability (H^2) was higher than 0.5 for all traits (Supplementary Table S2), demonstrative of the fact they were grown under tightly controlled environmental conditions. Flowering time, plant height and leaf area correlated positively with each other, and the content of all three carotenoids were highly correlated (Fig. 2).

Four unique QTL were identified in Wag18 across all traits

The Wag18 linkage map consisted of 279 markers covering 984.87 cM with an average chromosome length of 82.07 cM (Fig. 3a, b; Supplementary Table S3).

In Wag18, seven putative QTL (pQTL) were initially identified across all traits (significance less than 0.05 based on 1000 permutations, Supplementary Table S4). However, two pQTL for violaxanthin content, both on chromosome 1 (nearest markers at 5.95 Mb and 13.37 Mb) showed opposing additive effects despite similar genotype trends; individuals with the AB genotype exhibited significantly higher violaxanthin content than the AA (Malaysia-03) genotype at both loci ($p = 0.033$; $t = -2.15$; $df = 132.55$ and $p < 0.001$; $t = -3.43$; $df = 121.43$ respectively). The signal at *pVio2q* was driven by a small number of Malawi-02 genotypes (BB; $n = 20$) suggesting a spurious effect. Neither locus was retained in the final model following multiple QTL mapping, or single-scan validation and was thus excluded from further analysis.

Conversely, overlapping pQTL for plant height and leaf area showed consistent positive additive effects resulting from the Malawi-02 allele and mapped to a shared region on chromosome 1 around 89 cM (Table 1). Given the strong phenotypic correlation between these traits (Fig. 3), these signals were consolidated into a single robust QTL affecting overall plant size, hereafter referred to as *Siz1q*.

To summarise, extensive variation and high heritabilities were recorded for Wag18, and this enabled the mapping of four unique QTL associated with six agriculturally important traits (Table 1). *Siz1q* explained 25% and 18% of the phenotypic variance observed for plant height and leaf area respectively; *Lut1q* explained 10% of the lutein content variance; *Atc1q* contributed to 33% of the variance seen for alpha-tocopherol content; *Flt1q* explained 19% of the observed flowering time.

Phenotypic relationships in the Wag19 population

When the Wag19 founders were assessed, Malawi-01 had higher values than Malaysia-01 for carotenoid content, flowering time, plant height and leaf area, and a lower alpha-tocopherol content (Fig. 4), but the difference observed for all carotenoids and plant height was not statistically significant (Supplementary Table S5). For traits measured in Wag19 that were not assessed in Wag18, compared with Malaysia-01, vein density, trichome density, and stem colour was significantly higher in the Malawi-01 parent, while bundle sheath strand width and cell length were lower (Supplementary Table S5). Traits in Wag19 displayed transgressive segregation, except bundle sheath strand width and vein density, for which Malawi-02 demonstrated extremely low and high values respectively (Fig. 4).

Heritabilities were generally lower in the Wag19 population compared to Wag18, however only plant height (0.49) and bundle sheath cell length (0.19) had heritability of less than 0.5 (Supplementary Table S2). Flowering time correlated with leaf area in Wag19, but this was not the case for plant height (Fig. 5). Rather, in Wag19, plant height was also positively correlated with leaf area. Analysis of variance found that flowering time differed between stem colour categories ($p < 0.001$, $F = 10.07$, $df = 178$), with stems that lacked anthocyanin flowering earlier. As in Wag18, content of the carotenoids were highly correlated, however unlike in Wag18, a significant correlation was identified between alpha-tocopherol and carotenoid content (Fig. 4). Traits related to C_4 photosynthesis such as vein density, bundle sheath strand width, and bundle sheath cell length were strongly correlated, with the former having a negative relationship with the latter two traits. Interestingly, bundle sheath strand width was slightly, albeit significantly negatively correlated with carotenoid content. Plant height was positively correlated with bundle sheath size but negatively with vein density. A Chi-squared test revealed no significant relationship between stem colour and stem trichome density.

Increased marker density in Wag19 captured more cross-over events compared with Wag18

The Wag19 linkage map consisted of 920 markers covering 1645.02 cM with an average chromosome length of 126.54 cM (Fig. 6a, b;

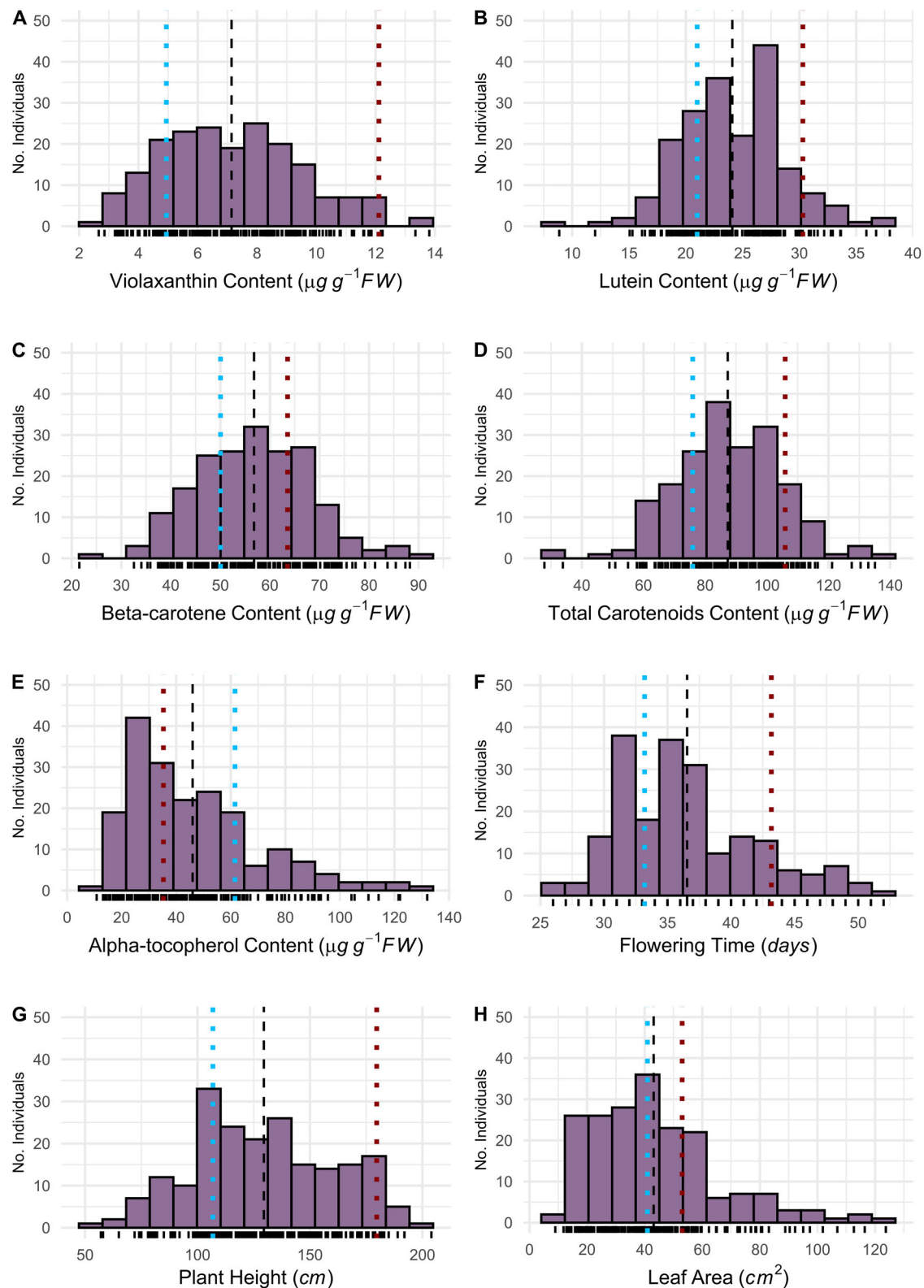


Fig. 1 | Transgressive segregation for all traits in the Wag18 F₂ mapping population. Blue and red dotted lines indicate means for the founders, Malaysia-03 and Malawi-02, respectively. The dashed black line represents is the mean of the F₂

population. **A** Violaxanthin content; **B** Lutein content; **C** Beta-carotene content; **D** Total carotenoid content; **E** Alpha-tocopherol content; **F** Flowering time; **G** Plant height; **H** Leaf area.

Supplementary Table S3. While the smallest four chromosomes, 3, 15, 16, and 17 (Hoang et al.²⁹) remained unmapped, the improved marker availability in Wag19, due to a less stringent selection process compared to Wag18 meant that chromosome 14 was captured in Wag19.

Eleven unique QTL were initially identified in the Wag19 population

In Wag19, a total of fourteen pQTL were identified (significance less than 0.05 based on 1000 permutations, Supplementary Table S6). Two pQTL

Fig. 2 | Correlation analysis for agronomic and nutrition traits in Wag18. Correlation coefficients were derived by Pearson correlation analysis and are presented in each box. Colour intensity indicates the strength of correlation, with deeper red representing a stronger positive relationship. The diagonal is the label for each phenotype. BSCW Bundle Sheath Cell Width. Significance is shown as *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

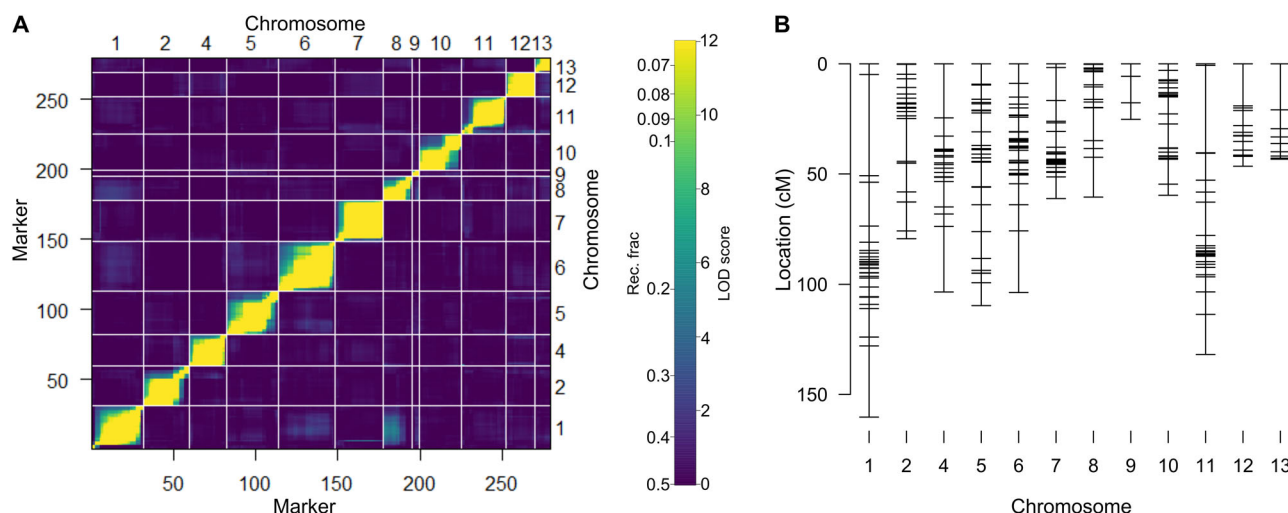
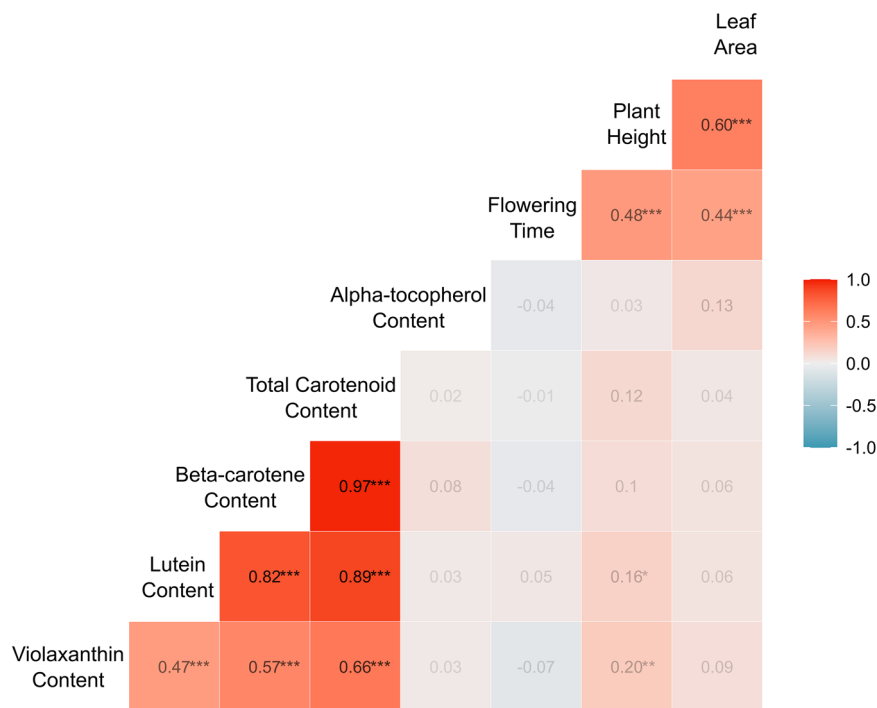


Fig. 3 | Wag18 genetic map. **A** Heatmap displaying recombination frequencies and LOD scores between each marker in the upper and lower triangles respectively; with yellow indicating a lower recombination frequency/higher LOD score, hence suggesting marker-pairs are linked. **B** Final genetic map.

were found for vein density, one on chromosome 4 at position 122.00 cM ($LOD = 3.88$) and one on chromosome 14 at 6.00 cM ($LOD = 4.30$; Supplementary Fig. S1a). Looking at the nearest markers for these two traits (4.59646348 and 14.00360911) it was clear the increase in vein density was, as expected, driven by the Malawi-01 (B) allele (Supplementary Fig. S1b, c), however 14.00360911 had just a single individual that was homozygous for the B allele. When this line was excluded only the loci on chromosome 4 remained significant in linear regression ($p < 0.001$). We therefore retained only this locus as a robust QTL and hereafter refer to it as *Vd1q*.

Two additional pQTL for plant height and leaf area were found on chromosome 1 at approximately the same location (84 cM). Both were positively influenced by the Malawi-01 (B) allele. As these traits were significantly correlated ($r = 0.33$; $p < 0.001$, Fig. 5), we consider these loci as a single multi-trait QTL affecting plant size, *Siz2q*.

For the discrete traits, stem trichome density and stem colour, we identified two pQTL each: *pStt1q* and *pStt2q* for trichomes, and *pStc1q* and

pStc2q for colour (Supplementary Table S7). Ordinal regression confirmed that both trichome-associated loci had significant effects ($AIC = 274.65$; Supplementary Table S7). For stem colour, the optimal model included both loci ($AIC = 160.91$), but only *pStc1q* was significant. A reduced model including only *pStc1q* had very similar AIC ($AIC = 161.57$), therefore we retained only these loci as a QTL for stem colour.

Finally, to ensure statistical robustness, all pQTL associated with stem trichome density and colour were cross-validated using non-parametric single-scan models. These yielded similar profiles for trichome density (Supplementary Fig. S2a), while the non-parametric scan for stem colour showed a slightly lower LOD peak (Supplementary Fig. S2b). Final LOD values reported for these traits are based on the non-parametric models.

In summary, eleven unique QTL affecting eight traits were identified in Wag19 (Table 2). One QTL was found to be significant for violaxanthin content (*Vio1q*), alpha-tocopherol content (*Atc2q*), vein density (*Vd1q*), flowering time (*Flt2q*), plant height (*Ht1q*), and stem colour (*Stc1q*); two

Table 1 | Overall summary of QTL identified in the Wag18 population

Phenotype	QTL number	QTL name	Genetic map position (chr:cM)	Nearest marker (chr.bp)	Physical position (Chr:Mb)	Physical range (Mb)	Additive effect	Dominance effect	LOD	Model formula	Model LOD	PVE by model (%)
Lutein content	Q1	<i>Lut1q</i>	2:18.0	2.2745371	2:9.34–27.87	18.53	−2.08	0.01	4.64	$y \sim Q1$	4.64	10.23
Alpha-tocopherol content ^a	Q1	<i>Atc1q</i>	9:5.7	9.00780456	9:0.00–1.32	1.32	−0.42	0.10	17.41	$y \sim Q1$	17.41	33.16
Flowering time	Q1	<i>Flt1q</i>	6:45.0	6.36025417	6:27.36–44.29	16.93	3.00	−2.02	9.19	$y \sim Q1$	9.19	19.24
Plant height	Q1	<i>Siz1q</i>	1:90.8	1.13687113	1:11.74–20.24	8.50	25.12	−1.30	12.38	$y \sim Q1$	12.38	25.02
Leaf area ^a	Q1	<i>Siz1q</i>	1:88.9	1.11928029	1:11.74–20.24	8.50	0.38	−0.11	8.33	$y \sim Q1$	8.33	18.38

Genetic map position is shown as "Chromosome:position (cM)", nearest marker name is based on physical position: "Chromosome:position (bp)". Negative and positive additive effects mean the allele from Malaysia-03 and Malawi-02 respectively is responsible for an increase in the trait.
PVE Percentage of Variance Explained.
^aAnalysis carried out on log transformed data. Numbers in italics were derived non-parametrically.

QTL were significantly associated with leaf area (*La1q* and *La2q*) and stem trichomes (*Stt1q* and *Stt2q*); and one QTL was found to effect overall plant size, i.e., having the same effect on plant height and leaf area (*Siz2q*). Overall, 12% of the variance observed for violaxanthin content, 14% for alpha-tocopherol content, 8% for vein density, 24% for flowering time, 25% for plant height, 29% for leaf area, 21% for stem trichomes and 52% for stem colour was explained by these QTL (Table 2). For traits assessed in both populations, four QTL were mapped in Wag18 (Fig. 7a) and seven were mapped in Wag19 (Fig. 7b).

Agriculturally important QTL found in both populations

In Wag18, the QTL *Siz1q* was located on chromosome 1 with a peak around 89.9 cM and in Wag19 the QTL *Siz2q* was located on chromosome 1 with a peak around 85.9 cM (Fig. 7). The nearest marker to both these QTLs were 1.13366757 and 1.17199903 (located at 113.37 Mb and 117.20 Mb) in Wag18 and Wag19 respectively and in both cases, the “African” allele (B) coming from Malawi-02 in Wag18 and Malawi-01 in Wag19, was responsible for the increase in plant size (Fig. 8a, b). The confidence limits based on the overlapping regions of the traits they were associated with (Fig. 7) for these QTL was between 87.7 cM and 90.7 cM for *Siz1q* and between 79.7 cM and 101.0 cM for *Siz2q*. Based on the nearest markers, the corresponding physical range for these QTL was 11.74–20.24 Mb and 11.15–61.79 Mb. While this region contains many genes, the overlapping QTL region means these effects may be under the control of the same gene or group of genes³⁸, and hereafter we will refer to this as single QTL, namely *Siz1q*.

Further, the QTL *Flt1q* (identified in Wag18) was located on chromosome 6 with a peak around 45.0 cM and *Flt2q* (identified in Wag19) was on chromosome 6 with a peak around 77.0 cM (Fig. 7). Once again, the African allele was responsible for the later flowering time (Fig. 8c, d). The confidence limits for these QTL were 39.2–54.5 cM and 54.1–86.0 cM corresponding to a physical range of 27.36–44.29 Mb and 36.68–49.41 Mb for *Flt1q* and *Flt2q* respectively. This overlap suggests the same underlying gene or group of genes may be responsible for this effect, from henceforth we shall refer to this as a single QTL, namely as *Flt1q*. Additionally, closer inspection of the nearest marker for the QTL for stem colour (*Stc1q*), found that just 2 out of 48 AA (Malaysia-01 genotype) individuals had a score of 2 (indicative of high anthocyanin) and no individuals with even one copy of the B (Malawi-01) allele were absent anthocyanin (a score of 0; Fig. 8e). Given its proximity to *Flt1q* (Fig. 7b), and the observation that stems with reduced anthocyanin flowered significantly earlier ($p < 0.001$, $F = 10.07$, $df = 178$), these loci may be genetically linked or share a common regulatory basis.

Discussion

Our identification of significant QTL in *G. gynandra* provides insight into the genetic basis of key agronomic and physiological traits, and paves the way for pre-breeding programmes. These findings not only contribute to the widespread adoption of *G. gynandra* as a food crop but also offer broader relevance for breeding nutrient-rich, climate resilient crops. Given that *G. gynandra* is currently harvested as a wild plant with minimal domestication³⁹, we anticipate these QTL will contribute to the development of improved germplasm, enhancing its potential for breeding efforts.

Two QTL were identified, *Siz1q* and *Flt1q*, that mapped to the same regions in both populations and could be the result of the same underlying gene(s); Lynch and Walsh, 1998). These controlled plant size and flowering time (Fig. 7). Interestingly, their effects were relatively stable in both populations, in the Wag18 population, *Siz1q* explained 25% and 18% of the variance associated with plant height and leaf area respectively, which was similar to the variance *Siz1q* explained by in Wag19 for these two traits, 17% and 10%, especially when considering the lower heritabilities associated with the Wag19 population (Supplementary Table S2). Further, *Flt1q* explained 19% and 24% of the variance associated with flowering time in Wag18 and Wag19 respectively. The recurrence of QTL in both populations suggests that key traits like plant size and flowering time are under conserved genetic control. This consistency across genetic backgrounds and environments is

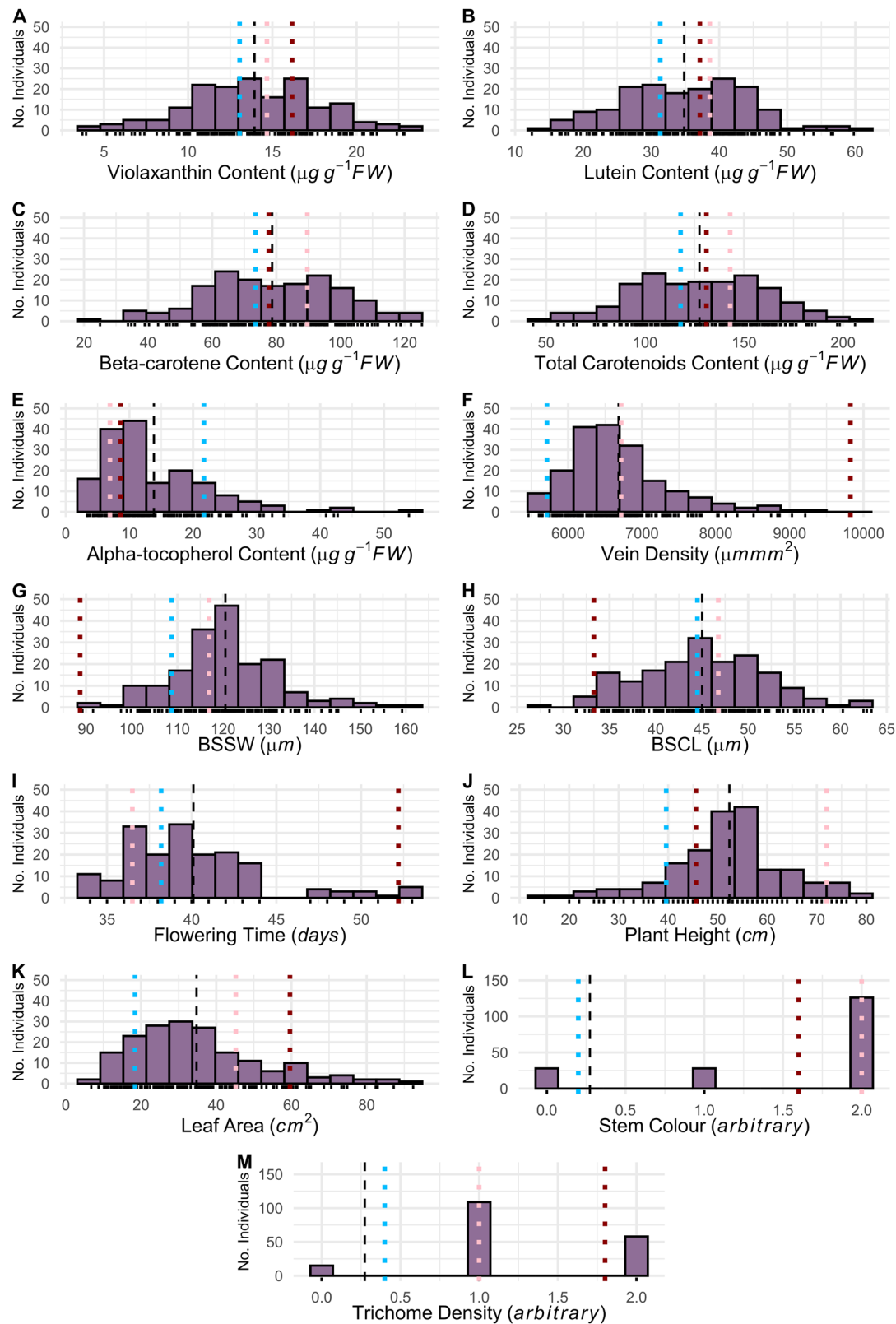


Fig. 4 | Phenotypic distributions in the Wag19 F₂ mapping population. The blue, red and pink dotted lines indicate the means for Malaysia-01, Malawi-02, and the F₁s respectively. The dashed black line represents the mean of the F₂ mapping population. **A** Violaxanthin content; **B** Lutein content; **C** Beta-carotene content; **D** Total carotenoids content; **E** Alpha-tocopherol content; **F** Vein density; **G** Bundle Sheath Strand Width (BSSW); **H** Bundle Sheath Cell Length (BSCL); **I** Flowering time; **J** Plant height; **K** Leaf area; **L** Stem colour; **M** Trichome density.

Fig. 5 | Correlation analysis for agronomic, nutrition and C₄ traits in the Wag19 population. Numbers indicate correlation coefficients derived by Pearson correlation analysis. Colour intensity indicates the strength of correlation, with deeper red representing a stronger positive relationship. The diagonal is the label for each phenotype. BSSW Bundle Sheath Strand Width, BSCL Bundle Sheath Cell Length (measured as length of 6 cells). Significance is shown as *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

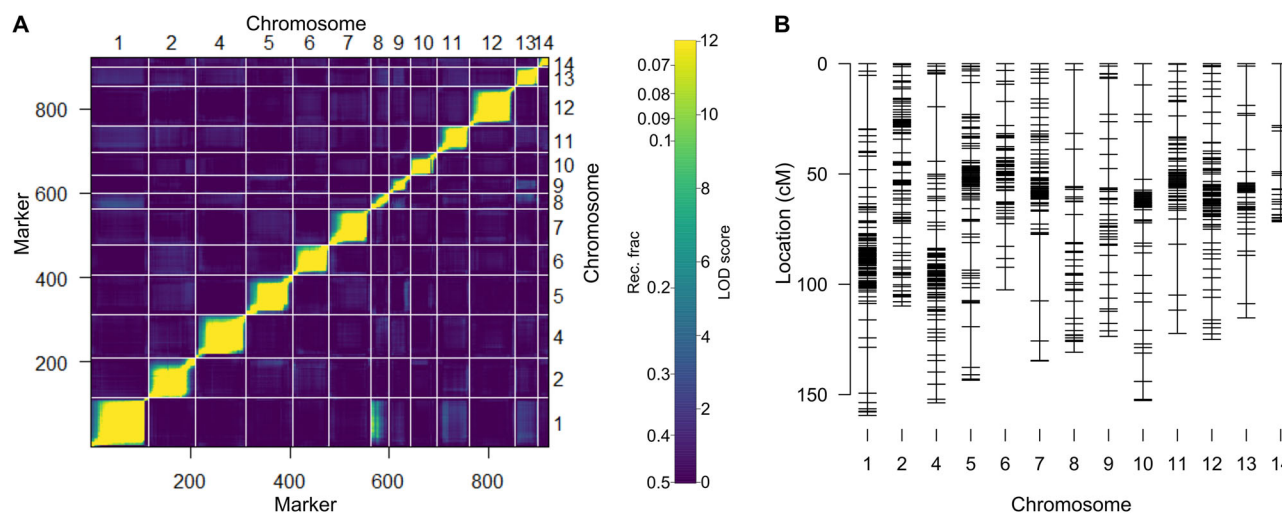
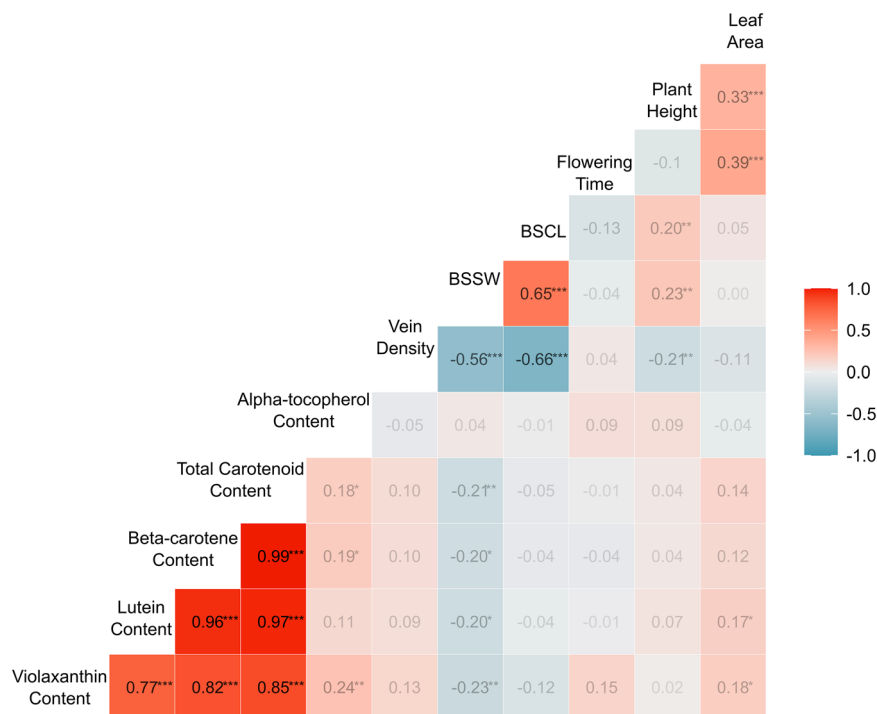


Fig. 6 | Wag19 genetic map. **A** Heatmap displaying recombination frequencies and LOD scores between each marker in the upper and lower triangles respectively; with yellow indicating a lower recombination frequency/higher LOD score, hence suggesting marker-pairs are linked. **B** Final genetic map.

encouraging for breeding programs, as it indicates the presence of stable, heritable targets.

Considering that *G. gynandra* is a commercially viable crop with nutritional and medical applications^{17,19–22}, identifying QTL associated with flowering time and plant size, two traits of value to plant breeders, in multiple environments and populations, and given their physical locations are known (Tables 1 and 2), means markers could be generated so desirable alleles could be incorporated into breeding germplasm via marker assisted selection⁴⁰. To our knowledge, these are the first example of stable targets for marker assisted selection identified in *G. gynandra* and could be of great benefit in the next stages of this plant's transcendence from underutilised to becoming a widely cultivated, climate resilient crop.

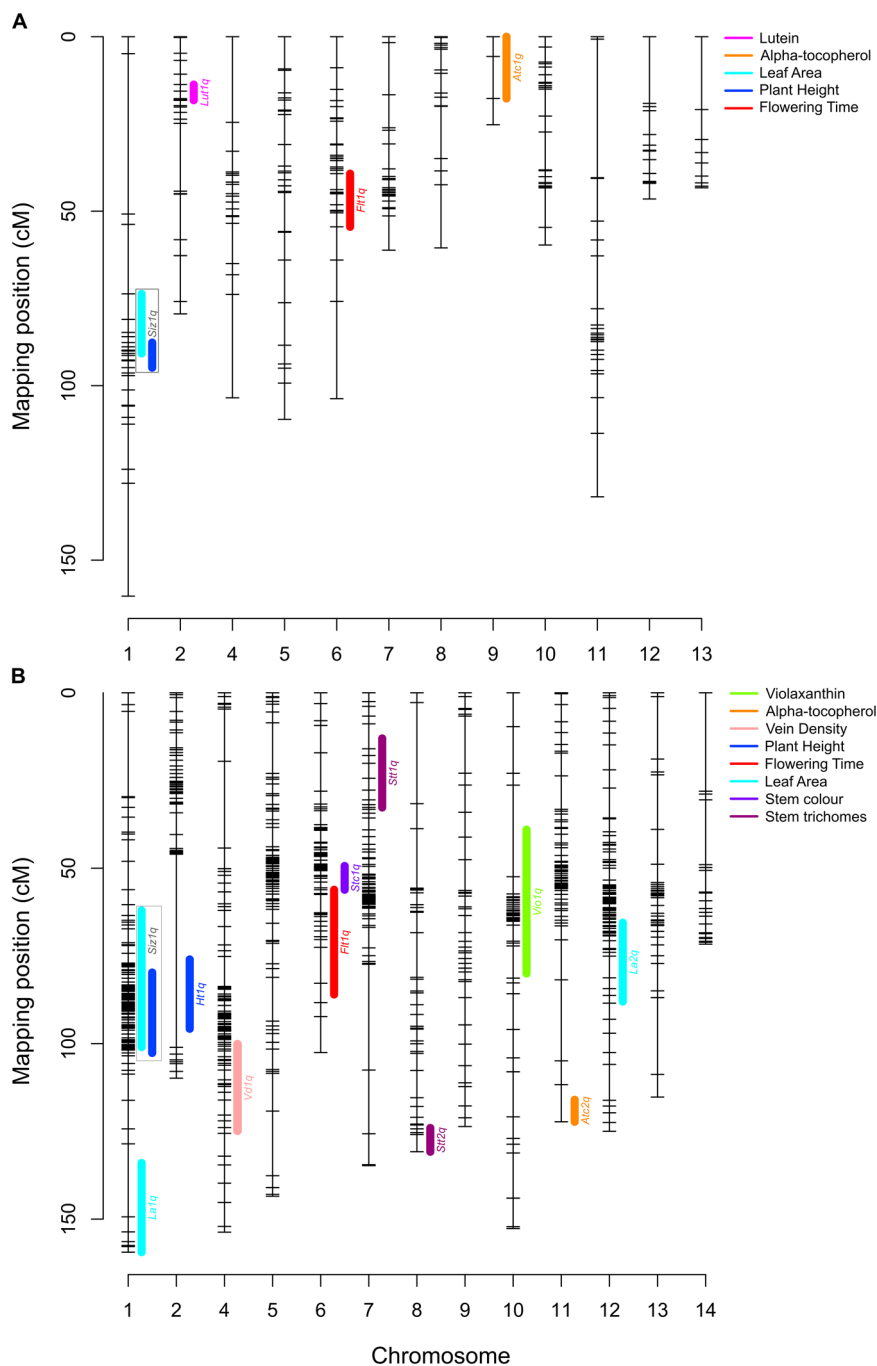
Further, across both populations, we identified four QTL associated with the content of the carotenoids, lutein and violaxanthin, and the vitamin E, alpha-tocopherol. The QTL for stem colour (*Stc1q*), also provides a

promising target, explaining over half of the observed variance. Its close proximity with the QTL for flowering time (*Flt1q*), suggests possible pleiotropy or linkage, and the early visibility of anthocyanin expression may offer a convenient phenotypic marker for breeders.

The vitamins we assessed are involved in protection of eyes from diseases, age-related degeneration^{41,42}, mitigating UV-induced skin damage⁴³ and are important for the immune system⁴⁴. There is a need for a targeted crop improvement programme for *G. gynandra*^{17,39} and these identified QTL could be useful in marker assisted selection for vitamin content. Furthermore, a previously reported negative correlation between carotenoids and tocopherols³¹ was not identified in the Wag18 population (Fig. 2), and was observed only weakly in the Wag19 population (Fig. 5) suggesting these two traits, can in fact be bred for simultaneously. A brief assessment of the F₁ lines in the Wag19 population also found strong heterosis for plant height (Fig. 4j). This has not been reported previously for

Table 2 Overall summary of QTL identified in the Wag19 population												
Phenotype	QTL number	QTL name	Genetic map position (chr:cM)	Nearest marker (chr:bp)	Physical location (Chr:Mb)	Physical range (Mb)	Additive effect	Dominance effect	LOD	Model formula	Model LOD	PVE by model (%)
Violaxanthin content	Q1	<i>Vio1q</i>	10:58.4	10.03027582	10:0.70–39.26	38.56	1.07	−2.20	4.77	$y \sim Q1$	4.77	12.26
Alpha-tocopherol content*	Q1	<i>Atc2q</i>	11:122.3	11.41198362	11:4.09–41.20	37.11	−0.23	−0.20	5.56	$y \sim Q1$	5.56	14.14
Vein density****	Q1	<i>Vd1q</i>	4:122.0	4.59646348	4:53.59–60.56	6.97	1.40	−0.36	3.88	$y \sim Q1$	3.88	8.75
Flowering time**	Q1	<i>Flt2q</i> (<i>Flt1q</i>)	6:77.0	6.46920704	6:36.68–49.41	12.73	0.083	−0.033	10.85	$y \sim Q1$	10.85	24.35
Plant height***	Q1	<i>Siz2q</i> (<i>Siz1q</i>)	1:83.5	1.17961377	1:11.15–61.79	50.64	514.58	419.87	7.46	$y \sim Q1 + Q2$	11.62	25.47
Plant height***	Q2	<i>Ht1q</i>	2:84.0	2.38898522	2:38.62–39.24	0.63	−528.14	12.25	4.48	$y \sim Q1 + Q2$	11.62	25.47
Leaf area****	Q1	<i>La1q</i>	1:145.0	1.70602349	1:70.06–70.92	0.85	−0.27	0.99	4.98	$y \sim Q1 + Q2 + Q3$	14.51	31.15
Leaf area****	Q2	<i>Siz2q</i> (<i>Siz1q</i>)	1:84.3	1.18700207	1:11.15–61.79	50.64	0.53	0.28	4.59	$y \sim Q1 + Q2 + Q3$	14.51	31.15
Leaf area****	Q3	<i>La2q</i>	12:73.0	12.38068542	12:2.70–39.64	36.94	0.73	−0.19	5.75	$y \sim Q1 + Q2 + Q3$	14.51	31.15
Stem trichomes	Q1	<i>Stt1q</i>	7:17.9	7.00721601	7:0.65–1.29	0.64	−0.26	−0.08	4.63	$y \sim Q1 + Q2$	8.87	20.10
Stem trichomes	Q2	<i>Stt2q</i>	8:129.0	8.44525256	8:44.28–44.53	0.25	0.25	0.15	4.24	$y \sim Q1 + Q2$	8.87	20.10
Stem colour	Q1	<i>Stc1q</i>	6:51.0	6.30828142	6:23.19–36.67	13.48	0.73	0.76	29.37	$y \sim Q1$	29.37	52.44
Genetic map position is shown as "Chromosome:position (cM)", nearest marker name is based on physical position: "Chromosome:position (bp)". Negative and positive additive effects mean the allele from Malaysia-01 and Malawi-01 respectively is responsible for an increase in the trait. Analysis carried out on *log, **cube-root, ***squared, ****square-root transformed data. Numbers in italics were derived non-parametrically.												
PVE Percentage of Variance Explained.												

Fig. 7 | QTL locations in both populations. Location of QTL in the (A) Wag18 and (B) Wag19 populations. QTL associated with each trait are colour-coded. Multi-trait QTL are represented in grey boxes. Confidence intervals were identified using the R/qtlTools package⁶⁷.



G. gynandra and indicates its suitability for hybrid breeding⁴⁵, which is of huge benefit for generating advanced breeding programmes¹⁶.

Cleomaceae is the closest related family to the Brassicaceae and therefore contains the most closely related C_4 species to the model C_3 plant *Arabidopsis thaliana* with which *G. gynandra* shares high similarity in gene sequence⁴⁶. This makes it an excellent model organism for the study of C_4 photosynthesis and it has been used as a system through which genetic determinants of C_4 enzymes have been investigated^{24–28}. C_4 photosynthesis has evolved independently over 65 times⁴⁷ and is more energy efficient under warm, dry conditions compared to the C_3 state. Mathematical modelling suggests that morphological changes such as increased vein density and larger bundle sheath cell size were crucial in the early evolution of the C_4 syndrome³⁷.

Kranz traits, vein density and bundle sheath size showed extensive variation (Fig. 4f–h) with one QTL, *Vd1q* being mapped to chromosome 4 (Fig. 7b). Further, initial QTL scans found an additional pQTL for vein

density (*pVd2q*; Supplementary Table S6). Although we deprioritised *pVd2q* as an outlier appeared to skew the results, rendering this pQTL association statistically insignificant after its removal, it is possible that this represents a genuine association. This possibility is supported by the additive effect observed in the heterozygote, and the large vein densities found in both ‘BB’ lines assessed (the Malawi-01 founder line and the F_2 line in question; Fig. 1f and Supplementary Fig. S1c). A population with a greater prevalence of these alleles would be needed to demonstrate this statistically.

Furthermore, it should be noted that in *G. gynandra* and other C_4 plants, carotenoids and tocopherols are crucial for protecting the photosynthetic apparatus from oxidative stress, ensuring efficient photosynthesis under intense light conditions and other abiotic stressors such as drought and high temperature^{48,49}, therefore the aforementioned QTL linked to these traits could also be useful for identifying lines suitable for growing under more extreme conditions.

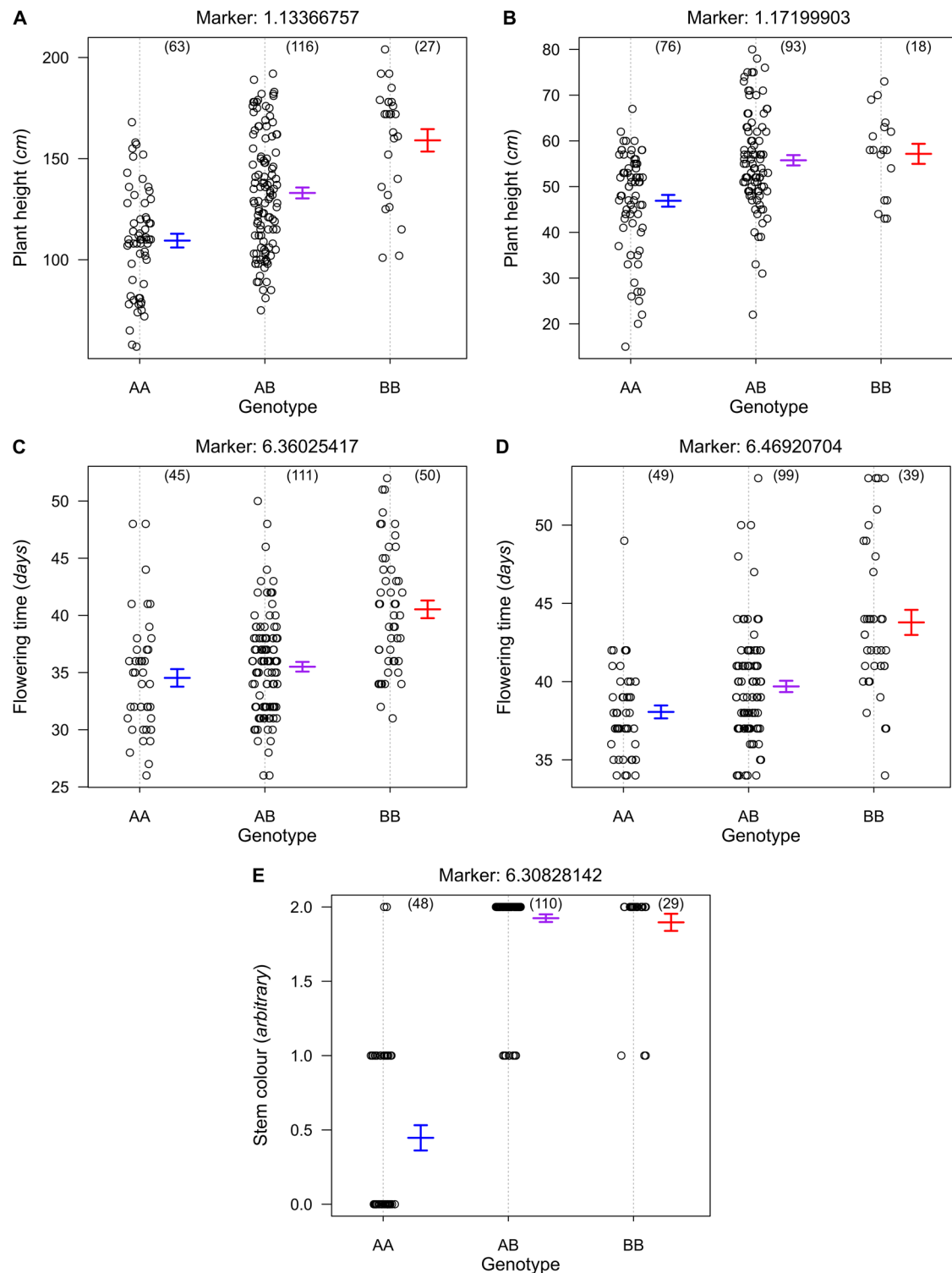


Fig. 8 | Phenotypes for agronomically important traits plotted against genotypes at markers nearest to associated QTL. **A** In Wag18, plant height (indicative of plant size) increased at the marker nearest to *Siz1q* when plants have the “African” allele; **B** the same trend was observed for Wag19. **C** Flowering time increased when plants had the B allele in the marker closest to *Flt1q* in Wag18 and **(D)** Wag19. **E** All

individuals homozygous for the Malaysia-01 (A) allele had no visible anthocyanin. Numbers in parentheses are the number of individuals with each genotype. Confidence intervals for average phenotypes within each genotype group are shown as blue for AA, purple for AB, and red for BB. Marker names describe physical location; “Chromosome:bp”.

To our knowledge, this is only the first example of linkage mapping being employed to identify QTL associated with C_4 characteristics in any dicotyledon. In maize (*Zea mays*) a bi-parental population identified QTL for stomatal size and density⁵⁰, and so together these data provide strong support for the notion that quantitative genetics could be used to study C_4

traits⁵¹. Given the region containing *Vdlq* contains 318 genes²⁹, we have refrained from reporting candidate genes due to the ambiguity in searching for GO terms amongst a relatively large number of genes. Instead, we aim here to prove that reverse genetics approaches such as QTL mapping is a viable option for the study of C_4 photosynthesis. Furthermore, alongside a

wild diversity panel, there is currently a Multi-Parental Advanced Generation Inter-Cross (MAGIC) population of *G. gynandra* being developed with the aim of mapping traits related to C_4 photosynthesis. Such association mapping methods are prone to false positives⁵² so populations derived from simple bi-parental crosses, such as that assessed here, that are equally related and therefore absent population structure, can add power to and complement such studies that enable higher resolution and the fine mapping of causal genes^{40,51}. We demonstrate here the effectiveness of mapping QTL in *G. gynandra* and, given its rapid generation time, phenotypic and genotypic diversity^{31,36}, combined with the power of bi-parental populations⁴⁰, it is an excellent resource for further mapping of additional traits. Furthermore, many F_1 hybrids already exist that can be used as a resource for future study of developmental, physiological or agricultural traits in *G. gynandra*.

Methods

Production of plant material, growth and phenotyping

Two populations of *G. gynandra* were developed at Wageningen University & Research in 2018 and 2019, hereafter referred to as Wag18 and Wag19. For Wag18, the female and male parents of the F_1 were derived from repeated self-fertilization of an early-flowering, short accession from Malaysia (TOT7200; Malaysia-03; MAY-03) and a late-flowering, tall accession from Malawi (TOT8917; Malawi-02; MAL-02; Sogbohossou et al.¹⁷). For Wag19, the female and male parents were derived from repeated self-fertilization of an early-flowering, short accession from Malaysia (TOT7199; Malaysia-01; MAY-01) and a late-flowering, short accession from Malawi (TOT8918; Malawi_01; MAL-01)¹⁷, these lines also have wide variation for C_4 photosynthetic traits³⁶. Wag18 consisted of 213 F_2 lines and Wag19 187 F_2 lines. As is standard for F_2 populations, each individual was genetically unique. Consequently, phenotypic replication at the level of individual genotypes was not possible. To increase confidence in QTL discovery despite this limitation we used two independent F_2 populations derived from divergent founder lines.

The populations were grown in the greenhouse of Wageningen University & Research, the Netherlands from March–June 2018 and April–July 2019. For Wag18 seven replicates of each parent were randomised, however, one replicate of Malaysia-03 was not healthy at the time of phenotyping and was therefore discarded. For Wag19 five replicates of each parent were randomised and two F_1 lines were included. Growing conditions were tightly controlled with plants grown under irrigated conditions with temperatures maintained at 24 °C in the day and 20 °C during night, and with artificial lights maintaining a minimum light intensity of 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and a photoperiod of 16 hr days and 8 hr nights. Sampling for DNA extraction and subsequent processing was carried out as described previously³⁰.

For both Wag18 and Wag19, plant height was measured after 10 weeks and the leaf area of three fully mature leaves per plant was assessed using image analysis in ImageJ⁵³. Flowering time was measured as the number of days from sowing to flowering. In Wag19, stem trichome density and stem colour were assessed as ordinal variables: 2 = dense trichomes or red colour; 1 = moderate trichome density and red/green colour; and 0 being no trichomes and green colour. Red colouration was indicative of anthocyanin content. Extraction and analysis of carotenoids and tocopherols as well as quantification was performed as described by³¹. In Wag18, seven plants died prior to vitamin-content assessment, along with an eighth plant that died before flowering. As a result, these plants were not assessed for plant height, leaf area and flowering time. Further, due to poor image quality or lack of replication, accurate leaf area could not be assessed for an additional nine plants. In Wag19, in total nineteen plants were not assessed for vitamin content due to sample contamination. Five plants died before being assessed for plant height, leaf area, stem colour, and stem trichome density. A further three died prior to flowering, and three plants could not be accurately assessed for leaf area. Lastly, six replicates of each parental line were assessed for vitamin content.

C_4 leaf anatomical traits were assessed in Wag19 as follows. Plants were harvested over a three-day period 4 weeks after germination. The most recent fully expanded leaf was selected for sampling. Tissue was harvested

from the central leaflet of each leaf and placed immediately in a plastic cuvette submerged in a 3:1 100% (v/v) ethanol:acetic acid fixative solution before treatment with 70% (v/v) ethanol solution overnight at 37 °C being refreshed one hour into treatment. Samples were cleared using 5% (w/v) NaOH solution for two hours at 37 °C before being washed with, and replaced in, 70% (v/v) ethanol until preparation for imaging. Immediately prior to imaging, samples were immersed in Lugol's solution (I_3K), and washed with water to highlight vein tissue as described by Simpson et al.⁵⁴. Samples were mounted with water and imaged on an Olympus BX41 light microscope with a mounted Micropublisher 3.3 RTV camera (Q Imaging). Images were captured with Q-Capture Pro 7 software. Bundle sheath cells were imaged at $\times 200$ magnification and assessed using the line measurement tool in ImageJ⁵³. Bundle sheath strand width was measured as the length of the bundle sheath strand perpendicular to the vein; six measurements were taken per field of view. Bundle sheath cell length was quantified by measuring the length parallel to the vein of 6 bundle sheath cells and dividing this by 6 to get a single value per field of view. For both parameters, six field of views were measured per plant. Four and five F_2 lines were not assessed for bundle sheath cell width and length respectively, in addition to one replicate of Malaysia-01 due to poor clarity of microscopy images. The Wag19 lines were also assessed for vein density, using six fields of view per plant at a magnification of $\times 100$ using the Starch4Kranz pipeline⁵⁴.

SNP calling

To call and select Single Nucleotide Polymorphisms (SNPs) in the two sets of founder lines for genotyping-by-sequencing in each population, raw reads generated for each founder³⁰ were checked for quality and trimmed using Trimmomatic⁵⁵. The reads were aligned to the reference genome using the BWA MEM algorithm⁵⁶ and Samtools⁵⁷. Duplicate reads were filtered and read groups added using GATK (version 4.0) Variant calling and filtering was performed with the HaplotypeCaller and the VariantFiltration tool of GATK⁵⁸. In order to select SNPs to be used for the characterization of the F_2 population, heterozygous SNPs and SNPs with no call were discarded. Non-polymorphic SNPs between both parents were removed using vcfilterjdk and JEXL expressions⁵⁹. SNP clusters in 150 bp windows were also removed. Furthermore, in the Wag18 founders, we subsequently used a set of re-sequenced genomes of 48 accessions: 24 from East and Southern Africa and 24 from Asia³⁰ to select SNPs that were present in at least 12 accessions from the same region of origin as the parents and, we identified SNPs that were present in genomic regions of interest for our targeted traits irrespective of whether they were unique to the parental lines or shared by other accessions from the same region. These additional filtering steps were not applied to the Wag19 founders. The final set consisted of 1309 SNPs and 9305 SNPs that were used for the genotyping-by-sequencing in the Wag18 and Wag19 population respectively.

Heritability estimates, linkage maps, and QTL mapping

Broad-sensed Heritability (H^2) was measured according to Kearsey and Pooni⁶⁰. The observed phenotypic variance (σ_p^2) is the sum of the genetic variance (σ_g^2) and environmental variance (σ_e^2). Since the homozygous parental lines are genetically identical, the observed variance within each line is equivalent to (σ_e^2) which can be extracted from the residual mean squares of the combined parental data. σ_g^2 can therefore be measured with the following equation, where σ_p^2 corresponds to the phenotypic variance excluding the parental lines:

$$\sigma_g^2 = \sigma_p^2 - \sigma_e^2 \quad (1)$$

This can then be used to calculate broad sensed heritability:

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} \quad (2)$$

Heritability estimates were carried out by comparing parental phenotypes to that of the final mapping population.

In both Wag18 and Wag19 populations, reads were aligned to the reference genome²⁹ using Bowtie2⁶¹ and Samtools⁵⁷. Haplotype-based variant detection was carried out using Freebayes⁶² and filtering was undertaken with Bcftools⁶³ before converting to R/qtl format using the R/utl package⁶⁴. To produce the final SNP set, markers with a Minor Allele Frequency (MAF) less than 0.2 were removed along with any markers missing in more than 70% of individuals and ambiguous genotypes (i.e., being heterozygous in the parents or homozygous in the F₁). From our initial sets of 1309 and 9305 SNPs, this resulted in final SNP sets of 521 and 2035 for the Wag18 and Wag19 populations respectively.

Linkage Maps were generated using the following pipeline. First, duplicate individuals were identified. In Wag18, three pairs of duplicate individuals were identified that shared more than 95% of their genotypes; each pair was next to each other in the experimental design suggesting a sampling error. Due to not knowing which individual was correct all 6 individuals were removed. No duplicates were identified in Wag19. Secondly, we removed one individual in the Wag18 population that was completely homozygous. We then constructed genetic maps as detailed by Broman and Sen⁶⁵ using the package R/qtl⁶⁶. Marker positions were based on their physical location²⁹, with errors being revealed through investigating marker recombination frequency, and successive use of the ‘droponemarker’ function⁶⁵. Local marker order was adjusted through rippling to minimize cross-over number by using a sliding window of 7 markers. Finally, individual genotyping errors were removed through study of double cross over rate and the ‘calc.errorlod’ function. Our final map for Wag18 consisted of 206 individuals, 279 markers, with 99.9% genotype rate, and an error rate of less than 0.001 based on loglikelihood scores. For Wag19, we attained the same error rate, 187 individuals and 920 markers, with a 99.7% genotype rate.

QTL were detected in both populations using the Multiple QTL Mapping (MQM) method described by Broman and Sen⁶⁵. First, single scans were performed using the ‘scanone’ function. Second, two-dimensional scans were carried out using the ‘scantwo’ function which considers QTL of large effect meaning modestly sized QTL can be identified in addition to interactions between QTL. Last, the putative QTL identified by ‘scanone’ and ‘scantwo’ scans were used as the starting model and the function ‘stepwiseqtl’ used to identify the final QTL model for each trait. For all scans, Haley-Knott regression for interval mapping was implemented for QTL detection due to its rapidity and that our genotype datasets had very little missing data. Further, if single scans revealed significant QTL not identified through MQM, these were investigated for significance through linear regression. Significance thresholds were calculated based on 1000 permutations per trait per scan. QTL models were fitted from the output of ‘stepwiseqtl’ and the applying the ‘fitqtl’ function, so that Percentage of Variance Explained (PVE) for each QTL model could be calculated as⁶⁵:

$$PVE = 1 - 10^{\left(\frac{-2 \times LOD}{n}\right)} \quad (3)$$

Where LOD is the Logarithm Of the Odds score associated with the QTL or QTL model in question and n is the number of individuals phenotyped. Confidence intervals for each QTL were calculated as the 1.5 LOD drop intervals based on the final QTL model and ‘scantwo’ permutations using the ‘qtlstats’ function from R/qtlTools⁶⁷. If the same QTL was identified for multiple traits, confidence limits were taken as being the overlapping region. Prior to mapping, phenotype data was transformed based on assessment of histograms in the mapping populations. Phenotype data were transformed when appropriate so that in the Wag18 population, leaf area and alpha-tocopherol-content underwent log-transformation, and in the Wag19 population, plant height was squared transformed, leaf area, and vein density were square-root transformed, flowering time was cube-root transformed, and alpha-tocopherol-content was log-transformed. To enable sufficient permutations to be generated for the non-parametric traits, stem trichome density and stem colour, normal models were selected while QTL mapping as it is accepted these still produce accurate results⁶⁵, however single scans were compared to the non-parametric model to ensure

QTL identified were correct. All QTL were identified under a significance threshold of 5% based on 1000 permutations derived from both the ‘scanone’ and ‘scantwo’ functions.

Statistical analysis

All statistics was carried out in R (version 4.1.3). Variation between parent lines was analysed using a two-sample Student’s *t* test in R Studio (V: 4.0.0). Data were checked for normality and equal variance in each group using Shapiro–Wilks test and Bartlett’s test respectively. If data were not normal in one or both groups, the Mann–Whitney U (MWU) test was used in place of the Student’s *t* test and variance was checked using Levene’s test. Results from Levene’s test determined comparisons made with the MWU test. If Levene’s test found equal variance, the MWU test for significance was based on median values. If Levene’s test found unequal variance between parental groups, the MWU test was based on variance. If data were normal in distribution but had unequal variance between groups, a Welch’s *t* test was used. For mapping populations, normality tests were determined from data distribution, and pairwise Pearson correlation analysis carried out using R/GGally package⁶⁸. To identify relationships between binary or discrete traits (stem trichome density, and stem colour) and continuous traits, ANOVA was used, while a Chi-squared test was used to test for a relationship between stem trichome density and stem colour. Linear regression was used to determine environmental effects and multiple QTL mapping performed as described. Plots were generated using R/ggplot2⁶⁹, R/corrplot⁷⁰, R/qtlTools⁶⁷, and R/qtl⁶⁶. Logistic regression was carried out using the R/MASS package⁷¹, and *p*-values derived from resulting *t*-distributions by applying the cumulative distribution function.

Data availability

Genotype and phenotype data can be found at <https://github.com/plycs5/GgQTL>. Sequencing datasets are available on request.

Received: 17 October 2024; Accepted: 19 May 2025;

Published online: 06 June 2025

References

1. FAO. *Dimensions of need: an atlas of food and agriculture* (FAO of the United Nations, 1995).
2. Tadele, Z. & Bartels, D. Orphan crops: their importance and the urgency of improvement. *Planta* **250**, 675–676 (2019).
3. Lara, S. W., Tsiame, A. & Cross, P. Discovering and mapping colloquial terminologies describing underutilized and neglected food crops - a comprehensive review. *Foods* **12**, 2428 (2023).
4. Fredenberg, E. et al. *Vision for adapted crops and soils (VACS) research in action opportunity crops for africa table of contents*, <https://www.rockefellerfoundation.org/wp-content/uploads/2024/03/Vision-for-Adapted-Crops-and-Soils-Research-in-Action-Summary-Report-Final.pdf> (2024).
5. Dwyer, W., Ibe, C. N. & Rhee, S. Y. Renaming Indigenous crops and addressing colonial bias in scientific language. *Trends Plant Sci.* **27**, 1189–1192 (2022).
6. Mabhaudhi, T. et al. Prospects of orphan crops in climate change. *Planta* **250**, 695–708 (2019).
7. FAO. *The state of food security and nutrition in the world 2021* (FAO, 2021).
8. Cassman, K. G., Grassini, P. & van Wart, J. Crop yield potential, yield trends, and global food security in a changing climate. In: *Handbook of climate change and agroecosystems* 37–51, https://doi.org/10.1142/9781848166561_0004 (Imperial College Press, 2010).
9. Cogato, A., Meggio, F., Migliorati, M. D. A. & Marinello, F. Extreme weather events in agriculture: a systematic review. *Sustainability* **11**, 1–18 (2019).
10. Dawson, I. K., Guarino, L. & Jaenicke, H. *Underutilised plant species: impacts of promotion on biodiversity: position paper No. 2*. (The International Centre for Underutilised Crops, 2007).

11. Tadele, Z. & Bartels, D. Promoting orphan crops research and development. *Planta* **250**, 675–676 (2019).
12. Hendre, P. S. et al. African Orphan Crops Consortium (AOCC): Status of developing genomic resources for African orphan crops. *Planta* **250**, 989–1003 (2019).
13. FAO. *Rediscovering hidden treasures of neglected and underutilized species for zero hunger in Asia* (FAO of the United Nations, 2018).
14. Talabi, A. O. et al. Orphan crops: a best fit for dietary enrichment and diversification in highly deteriorated marginal environments. *Front. Plant Sci.* **13**, 839704 (2022).
15. Alandia, G., Rodriguez, J. P., Jacobsen, S. E., Bazile, D. & Condori, B. Global expansion of quinoa and challenges for the Andean region. *Glob. Food Sec.* **26**, 100429 (2020).
16. Yadav, O. P. et al. Genetic gains in pearl millet in India: insights into historic breeding strategies and future perspective. *Front. Plant Sci.* **12**, 645038 (2021).
17. Sogbohossou, E. O. D. et al. A roadmap for breeding orphan leafy vegetable species: a case study of *Gynandropsis gynandra* (Cleomaceae). *Hortic. Res.* **5**, 2 (2018).
18. van Zonneveld, M. et al. Forgotten food crops in sub-Saharan Africa for healthy diets in a changing climate. *Proc. Natl. Acad. Sci.* **120**, e2205794120 (2023).
19. Moyo, M. et al. Determination of mineral constituents, phytochemicals and antioxidant qualities of *Cleome gynandra*, compared to *Brassica oleracea* and *Beta vulgaris*. *Front. Chem.* **5**, 128 (2018).
20. Chweya, J. A. & Mnzava, N. A. *Cat's Whiskers. Cleome gynandra L. promoting the conservation and use of underutilized and neglected crops*. 11. (Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, 1997).
21. Omondi, E. O. et al. Mating biology, nuclear DNA content and genetic diversity in spider plant (*Cleome gynandra*) germplasm from various African countries. *Plant Breed* **136**, 578–589 (2017).
22. Nijveldt, R. J. et al. Flavonoids: A review of probable mechanisms of action and potential applications. *Am. J. Clin. Nutr.* **74**, 418–425 (2001).
23. Muchuweti, M., Mupure, C., Ndhala, A., Murenje, T. & Benhura, M. A. N. Screening of antioxidant and radical scavenging activity of *Vigna unguiculata*, *Bidens pilosa* and *Cleome gynandra*. *Am. J. Food Technol.* **2**, 161–168 (2007).
24. Brown, N. J. et al. Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. *Science* **331**, 1436–1439 (2011).
25. Kajala, K. et al. Multiple *Arabidopsis* genes primed for recruitment into C4 photosynthesis. *Plant J* **69**, 47–56 (2012).
26. Williams, B. P. et al. An untranslated cis-element regulates the accumulation of multiple C4 enzymes in *Gynandropsis gynandra* mesophyll cells. *Plant Cell* **28**, 454–465 (2016).
27. Reyna-Llorens, I. et al. Ancient duons may underpin spatial patterning of gene expression in C4 leaves. *Proc. Natl. Acad. Sci. USA* **115**, 1931–1936 (2018).
28. Singh, P. et al. C4 gene induction during de-etiolation evolved through changes in cis to allow integration with ancestral C3 gene regulatory networks. *Sci. Adv.* **9**, eade9756 (2023).
29. Hoang, N. V. et al. The *Gynandropsis gynandra* genome provides insights into whole-genome duplications and the evolution of C4 photosynthesis in Cleomaceae. *Plant Cell* **35**, 1334–1359 (2023).
30. Sogbohossou, E. O. D. Orphan No More? Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae). <https://doi.org/10.1126/science.333.6040.268-a> (2019).
31. Sogbohossou, E. O. D. et al. Association between vitamin content, plant morphology and geographical origin in a worldwide collection of the orphan crop *Gynandropsis gynandra* (Cleomaceae). *Planta* **250**, 933–947 (2019).
32. Kiebre, Z., Bationo/Kando, P., Sawadogo, N., Sawadogo, M. & Zongo, J. Selection of phenotypic interests for the cultivation of the plant *Cleome gynandra* L. in the vegetable gardens in Burkina Faso. *J. Exp. Biol. Agric. Sci.* **3**, 288–297 (2015).
33. Acquaah, G. *Principles of plant genetics and breeding* (Wiley, 2012). <https://doi.org/10.1017/CBO9781107415324.004>.
34. Flood, J., Harbinson, J. & Aarts, M. G. M. Natural genetic variation in plant photosynthesis. *Trends Plant Sci.* **16**, 327–335 (2011).
35. Wright, T. I. C. et al. Identification of quantitative trait loci relating to flowering time, flag leaf and awn characteristics in a novel *Triticum dicoccum* mapping population. *Plants* **9**, 829 (2020).
36. Reeves, G. et al. Natural variation within a species for traits underpinning C4 photosynthesis. *Plant Physiol.* **177**, 504–512 (2018).
37. Williams, B. P., Johnston, I. G., Covshoff, S. & Hibberd, J. M. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *Elife* **2**, e00961 (2013).
38. Lynch, M. & Walsh, B. *Genetics and analysis of quantitative traits* (Sinauer Associates, 1998). <https://doi.org/10.1086/318209>.
39. Chataika, B. et al. Diversity and domestication status of Spider Plant (*Gynandropsis gynandra*, L.) amongst sociolinguistic groups of northern Namibia. *Agronomy* **10**, 56 (2020).
40. Cockram, J. & Mackay, I. Genetic mapping populations for conducting high-resolution trait mapping in plants. *Adv. Biochem. Eng. Biotechnol.* **164**, 109–138 (2018).
41. Cazzonelli, C. I. Carotenoids in nature: Insights from plants and beyond. *Funct. Plant Biol.* **38**, 833–847 (2011).
42. Esteban, R., Moran, J. F., Becerril, J. M. & García-Plazaola, J. I. Versatility of carotenoids: an integrated view on diversity, evolution, functional roles and environmental interactions. *Environ. Exp. Bot.* **119**, 63–75 (2015).
43. Aguilera, J., de Gálvez, M. V., Sánchez, C. & Herrera-Ceballos, E. Changes in photoinduced cutaneous erythema with topical application of a combination of vitamins C and E before and after UV exposure. *J. Dermatol. Sci.* **66**, 216–220 (2012).
44. Lewis, E. D., Meydani, S. N. & Wu, D. Regulatory role of vitamin E in the immune system and inflammation. *IUBMB Life* **71**, 487–494 (2019).
45. Mackay, I. J., Cockram, J., Howell, P. & Powell, W. Understanding the classics: the unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotechnol. J.* **19**, 26–34 (2021).
46. Bräutigam, A. et al. An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. *Plant Physiol* **155**, 142–156 (2011).
47. Sage, R. F., Christin, P.-A. & Edwards, E. J. The C4 plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169 (2011).
48. Young, A. J. The photoprotective role of carotenoids in higher plants. *Physiol. Plant.* **83**, 702–708 (1991).
49. Hasanuzzaman, M., Nahar, K. & Fujita, M. Role of tocopherol (Vitamin E) in plants: abiotic stress tolerance and beyond. In: *Emerging technologies and management of crop stress tolerance* vol. 2, 267–289 (Academic Press, 2014).
50. Xie, J. et al. Optical topometry and machine learning to rapidly phenotype stomatal patterning traits for maize QTL mapping. *Plant Physiol.* 1–19, <https://doi.org/10.1093/plphys/kiab299> (2021).
51. Simpson, C. J. C., Reeves, G., Tripathi, A., Singh, P. & Hibberd, J. M. Using breeding and quantitative genetics to understand the C4 pathway. *J. Exp. Bot.* **73**, 3072–3084 (2022).
52. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29 (2013).
53. Schneider, C., Rasband, W. & Eliceiri, K. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
54. Simpson, C. J. C., Singh, P., Sogbohossou, D. E. O., Eric Schranz, M. & Hibberd, J. M. A rapid method to quantify vein density in C4 plants using starch staining. *Plant Cell Environ.* **46**, 2928–2938 (2023).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

56. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <http://arxiv.org/abs/1303.3997> (2013).
57. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **11**, 11.10.1–11.10.33 (2014).
59. Lindenbaum, P. & Redon, R. Bioalcaide, samjcs and vcfilterjcs: Object-oriented formatters and filters for bioinformatics files. *Bioinformatics* **34**, 1224–1225 (2018).
60. Kearsey, M. J. & Pooni, H. S. *The genetical analysis of quantitative traits* (Springer, 1996).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
62. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <http://arxiv.org/abs/1207.3907> (2012).
63. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
64. Walsh, T. & Hu, Y. *_util: Utility functions for R/qtl analysis*. R package version 0.2.0, commit 1949e1d3159b453bcf1615a35a406064116943bc, <https://github.com/gact/util> (2020).
65. Broman, K. & Sen, S. *A guide to QTL mapping with R/qtl* (Springer, 2009). <https://doi.org/10.18637/jss.v032.b05>
66. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
67. Lovell, J. T. *_qtlTools: data processing and plotting in association with R/qtl*. R package version 1.2.0, commit 96f6b61e255314f6a5a32e38105c160dc52c037a, <https://github.com/jtlovell/qtlTools> (2018).
68. Schloerke B, et al. *_GGally: Extension to 'ggplot2'_*. R package version 2.2.1, <https://CRAN.R-project.org/package=GGally> (2024).
69. Wickham, H. & Chang, W. *ggplot2: an Implementation of the Grammar of Graphics*. GNU General Public License, <https://cran.r-project.org/web/packages/ggplot2> (Cran, 2007).
70. Taiyun Wei and Viliam Simko. R package 'corrplot': Visualization of a Correlation Matrix (Version 0.95) <https://github.com/taiyun/corrplot> (2024).
71. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).

Acknowledgements

The work was supported by a BBSRC DTP studentship to Conor J. C. Simpson, European Research Council Grant 694733 Revolution to Julian M. Hibberd, and Netherlands Organization for Scientific Research Grant W.08.270.350 to M. Eric Schranz. P.S. was supported by the UK Research and Innovation Future Leaders Fellowship (MR/Y016882/1); REVOLUTON,

awarded to P.S. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

Author contributions

C.J.C.S., P.S., D.S., E.S., and J.M.H. conceptualised and designed the study. C.J.C.S., D.S., and P.S. collected data. D.S. performed extracts and made the crosses from which Wag18 was derived. G.R. and P.S. made the crosses from which Wag19 was derived. D.S. called SNPs for genotyping-by-sequencing. C.J.C.S. mapped SNPs to the reference genome. C.J.C.S. carried out microscopy, heritability and QTL analysis. C.J.C.S. generated the figures. C.J.C.S., P.S., J.M.H., and E.S. wrote the article. J.M.H., P.S. and E.S. oversaw and supervised this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44264-025-00074-0>.

Correspondence and requests for materials should be addressed to Pallavi Singh or Julian M. Hibberd.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025